Introduction to Complex Networks

Jesús Gómez Gardeñes

Departamento de Física de la Materia Condensada

Instituto de Biocomputación y Física de Sistemas Complejos (BIFI)

Universidad de Zaragoza

Complex Networks in a Nutshell

Jesús Gómez Gardeñes

Departamento de Física de la Materia Condensada
Instituto de Biocomputación
y Física de Sistemas Complejos (BIFI)

Universidad de Zaragoza

Complex Systems

Ingredients

- Composed of many **interacting** elements
- They give rise to **emergent** collective behavior

Complex Systems

Ingredients

Composed of many **interacting** elements

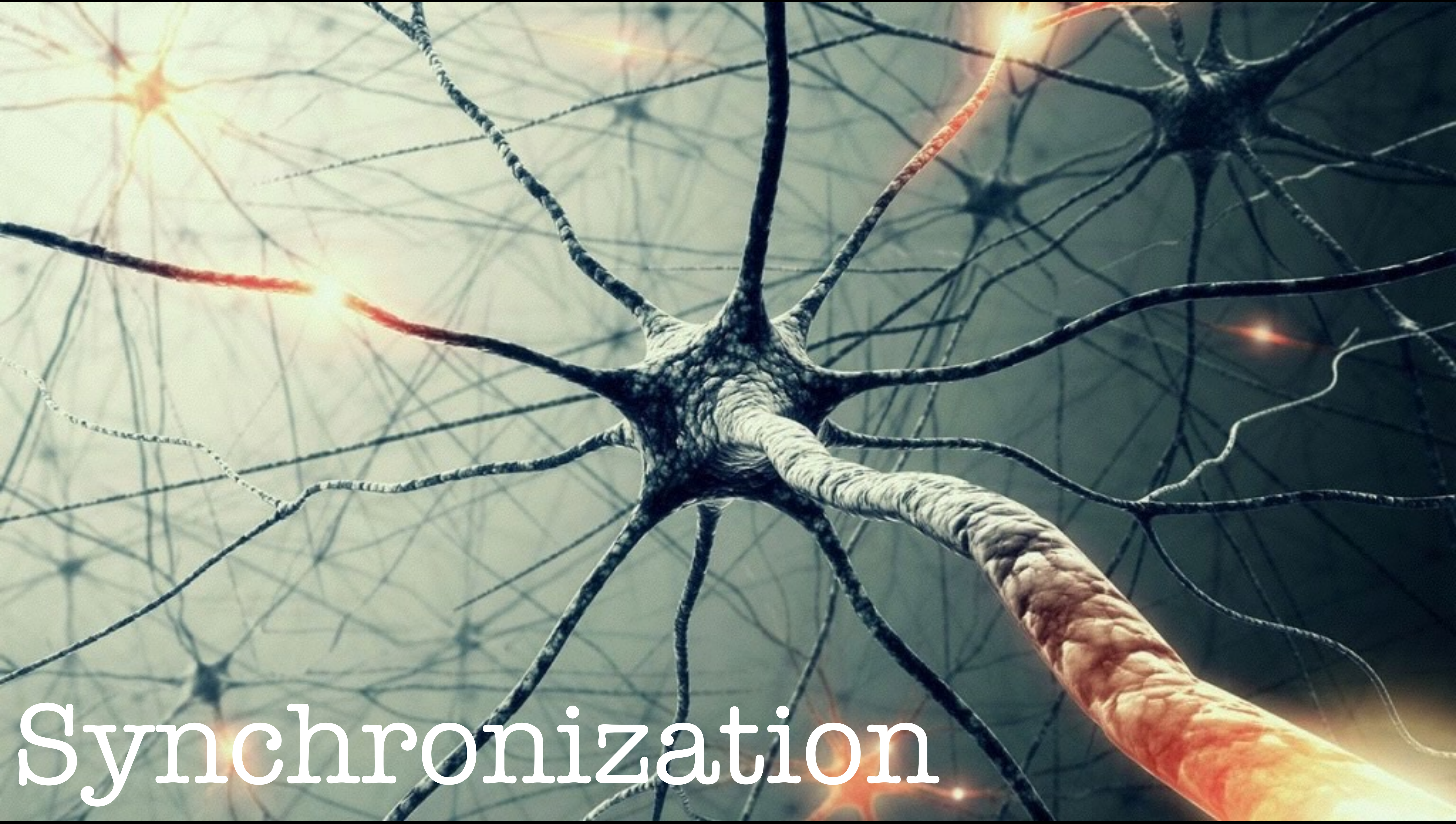
They give rise to **emergent** collective behavior

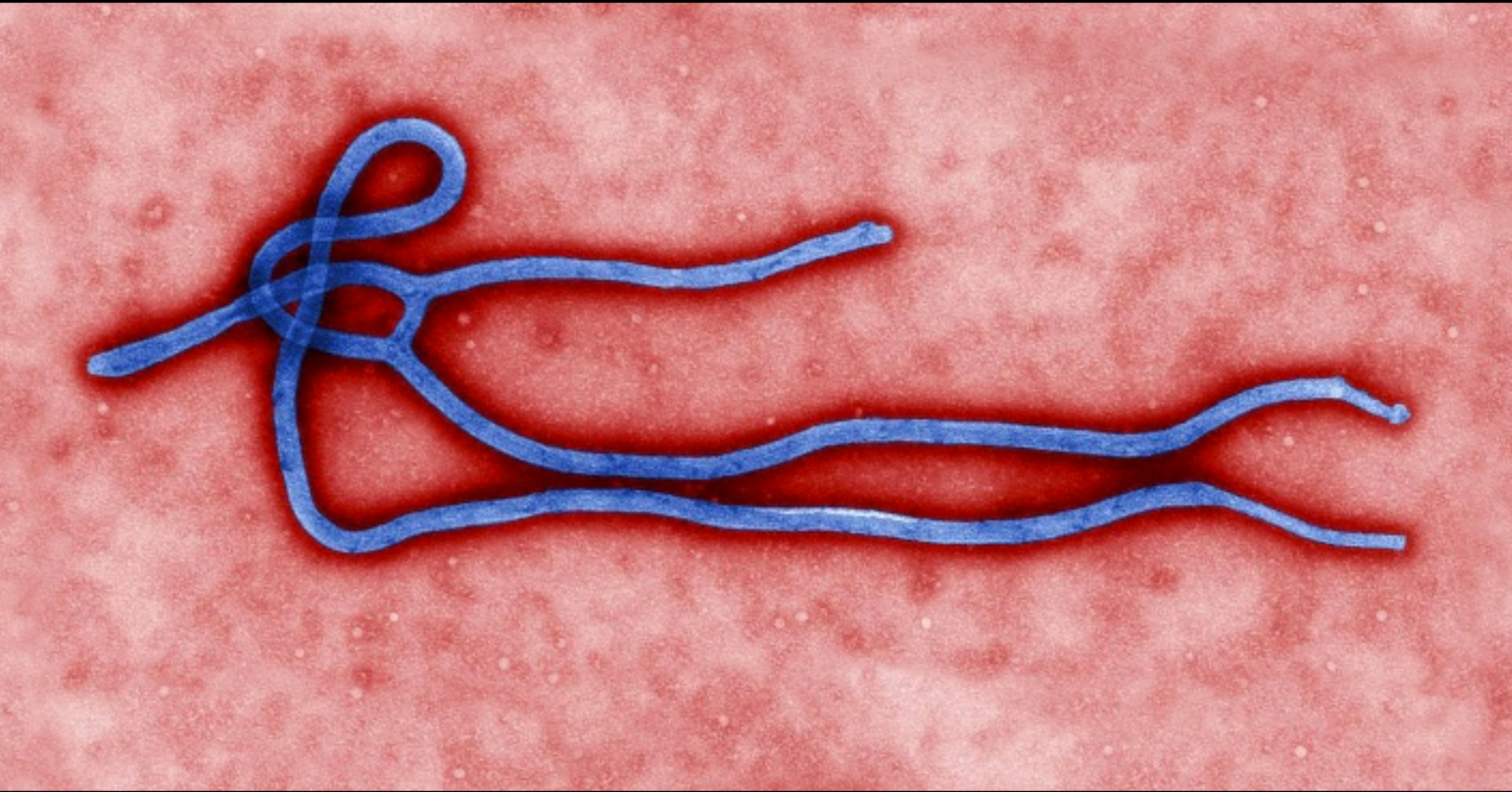
Emergence: Not directly related to individual properties

They are **Ubiquitous**, *i.e.*, not related to any characteristic life/energy scale

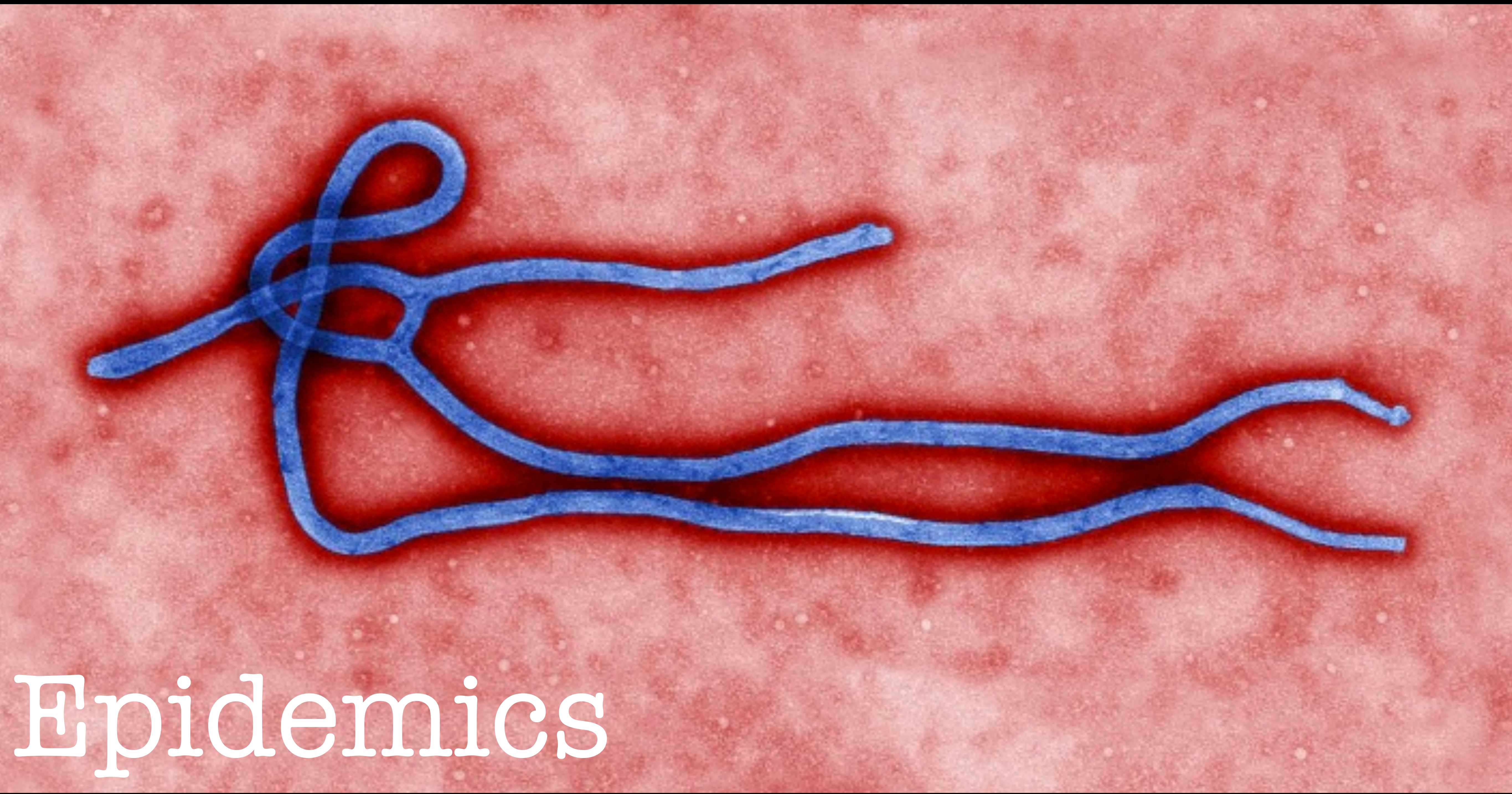
@gomezgardenes

@gomezgardenes


Synchronization



@gomezgardenes


Epidemics

@gomezgardenes

@gomezgardenes

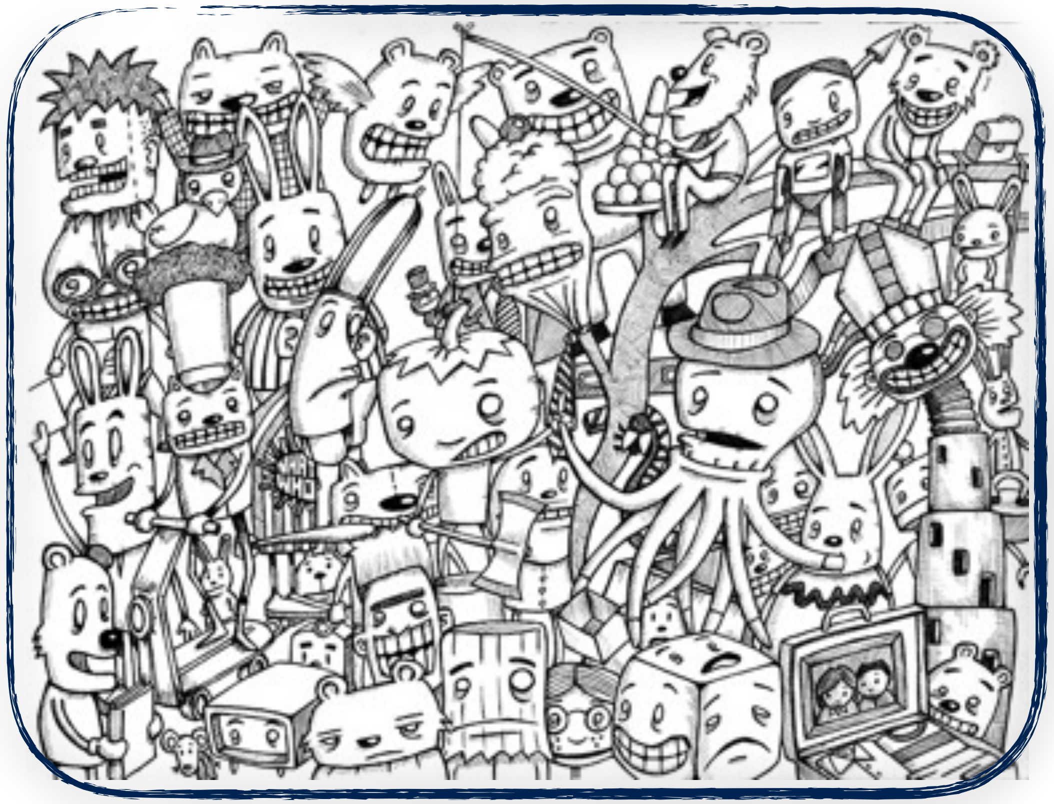

@gomezgardenes

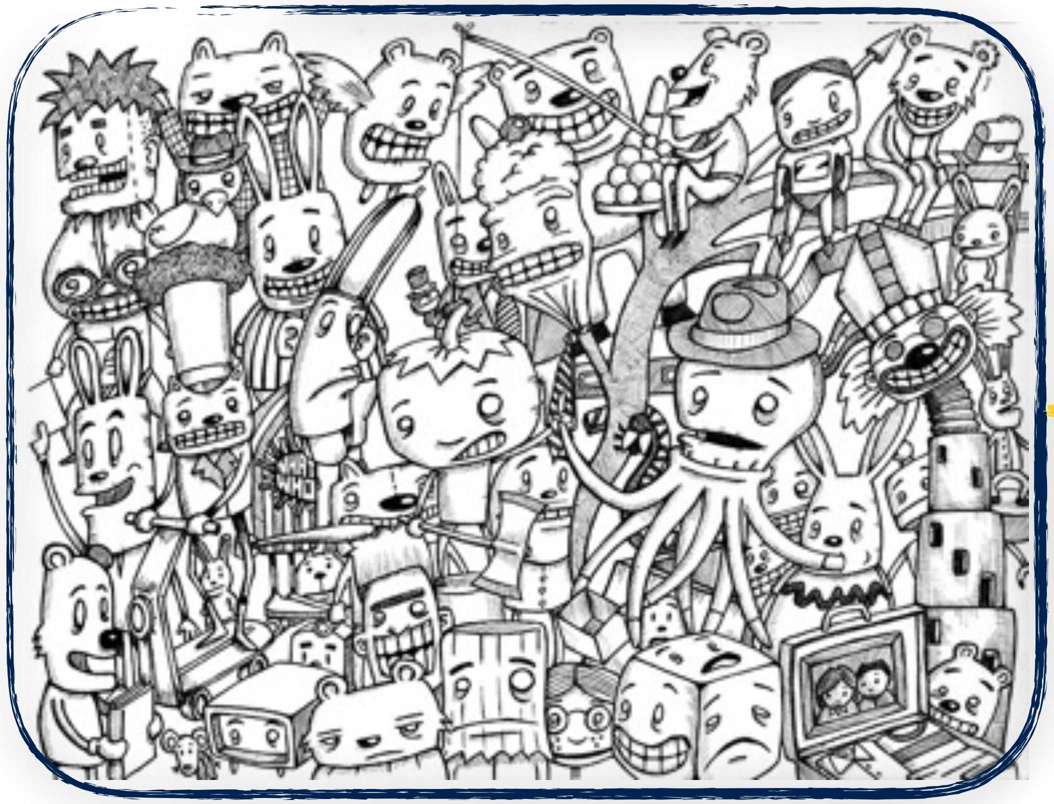
more than Congestion

Social Collective Behavior

- Social Norms & Conventions
- Economic Crisis
- Viral Information
- Unfolding of social movements
- …

@gomezgardenes





+ Realism

+ Complexity

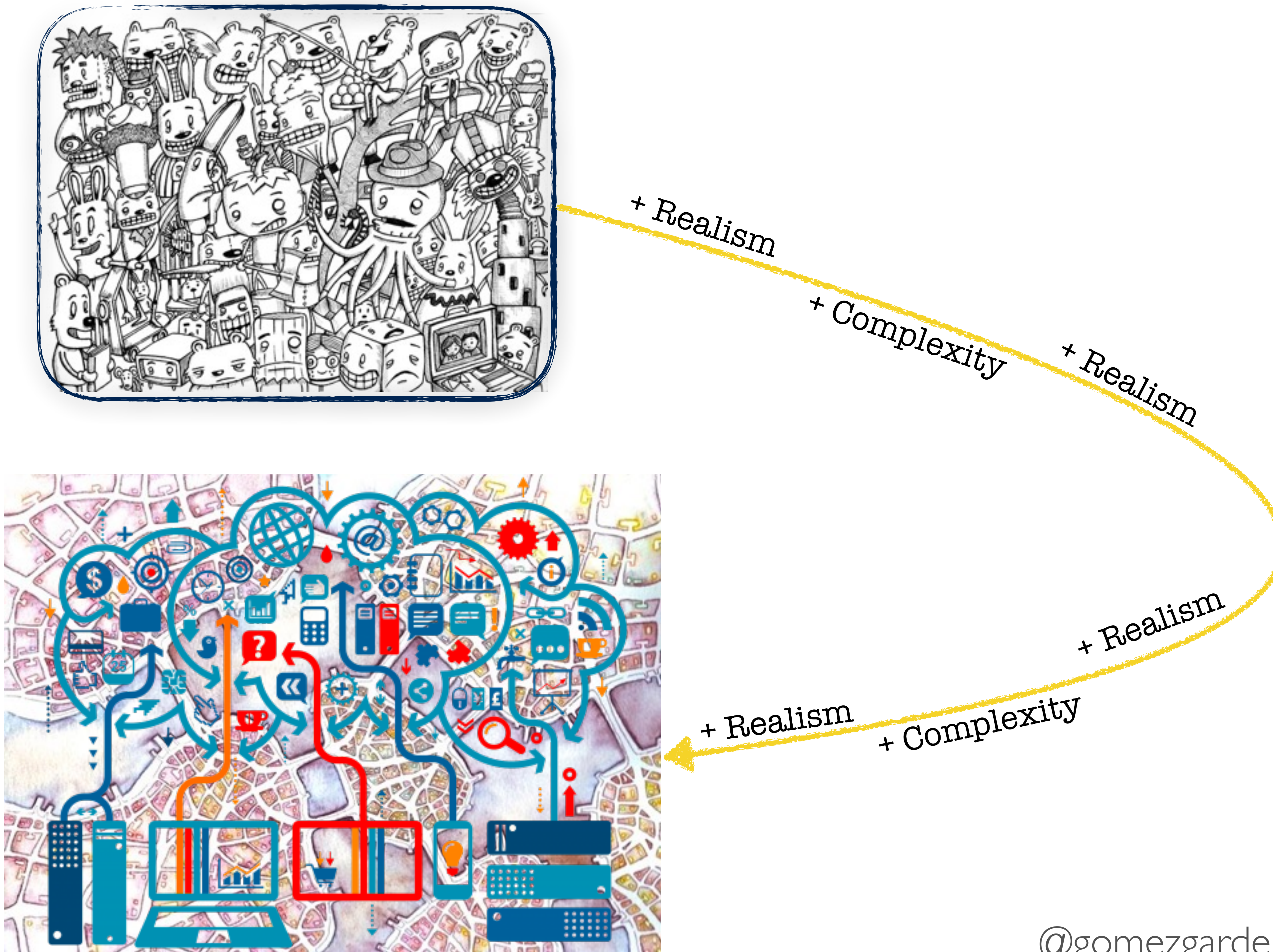
+ Realism

+ Realism

+ Complexity

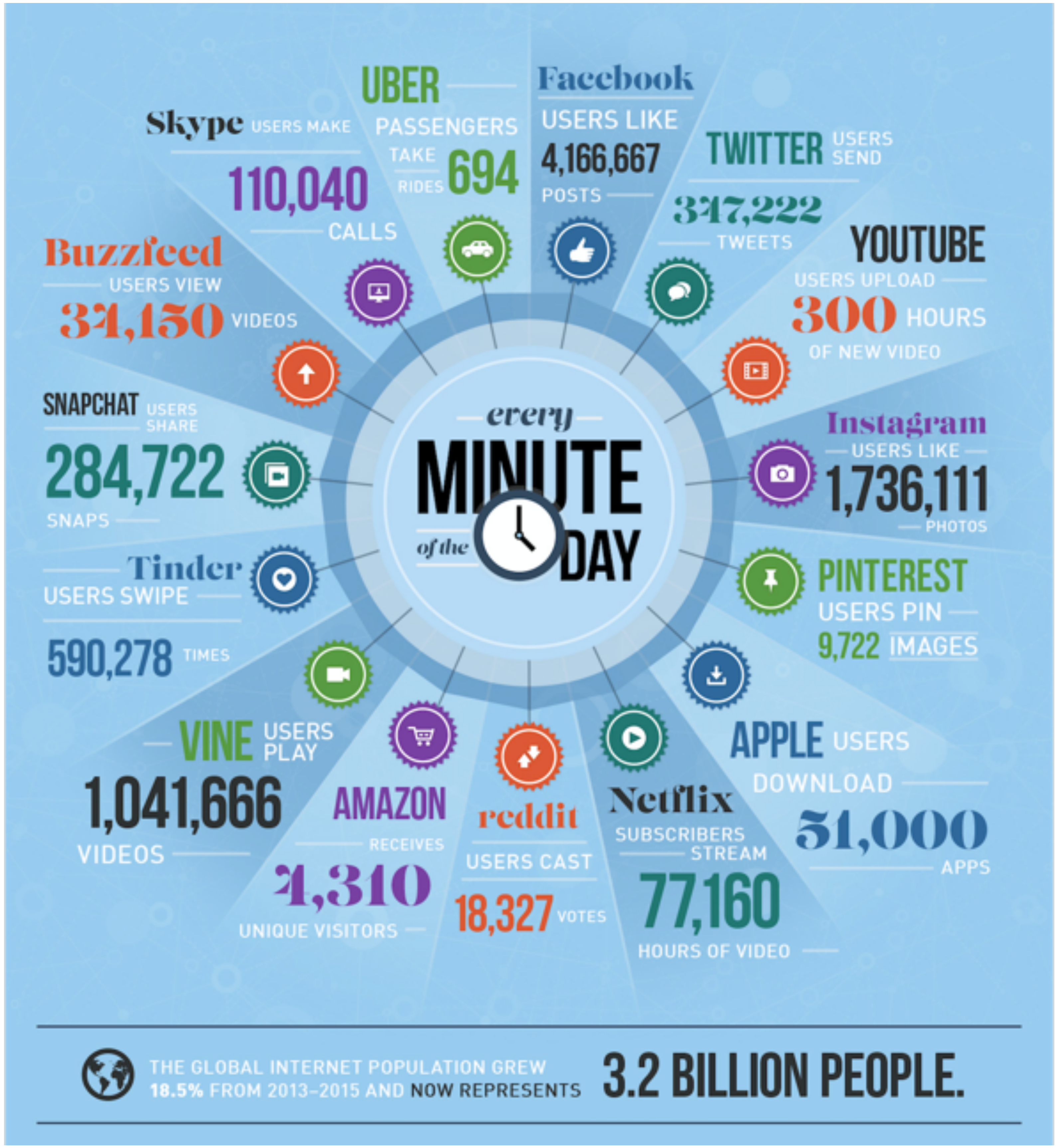
+ Realism

@gomezgardenes



@gomezgardenes

BIG DATA Opportunities



@gomezgardenes

Use
&
Misuse

LANGUAGE USE

OPEN ACCESS Freely available online

PLOS | ONE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

Delia Mocanu[1], Andrea Baronchelli[1], Nicola Perra[1], Bruno Gonçalves[2], Qian Zhang[1], Alessandro Vespignani[1,3,4]*

1 Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, Massachusetts, United States of America, **2** Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France, **3** Institute for Quantitative Social Sciences at Harvard University, Cambridge, Massachusetts, United States of America, **4** Institute for Scientific Interchange Foundation, Turin, Italy

Abstract

Large scale analysis and statistics of socio-technical systems that just a few short years ago would have required the use of consistent economic and human resources can nowadays be conveniently performed by mining the enormous amount of digital data produced by human activities. Although a characterization of several aspects of our societies is emerging from the data revolution, a number of questions concerning the reliability and the biases inherent to the big data "proxies" of social life are still open. Here, we survey worldwide linguistic indicators and trends through the analysis of a large-scale dataset of microblogging posts. We show that available data allow for the study of language geography at scales ranging from country-level aggregation to specific city neighborhoods. The high resolution and coverage of the data allows us to investigate different indicators such as the linguistic homogeneity of different countries, the touristic seasonal patterns within countries and the geographical distribution of different languages in multilingual regions. This work highlights the potential of geolocalized studies of open data sources to improve current analysis and develop indicators for major social phenomena in specific communities.

Citation: Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981

Editor: Yamir Moreno, University of Zaragoza, Spain

Received January 5, 2013; **Accepted** March 18, 2013; **Published** April 18, 2013

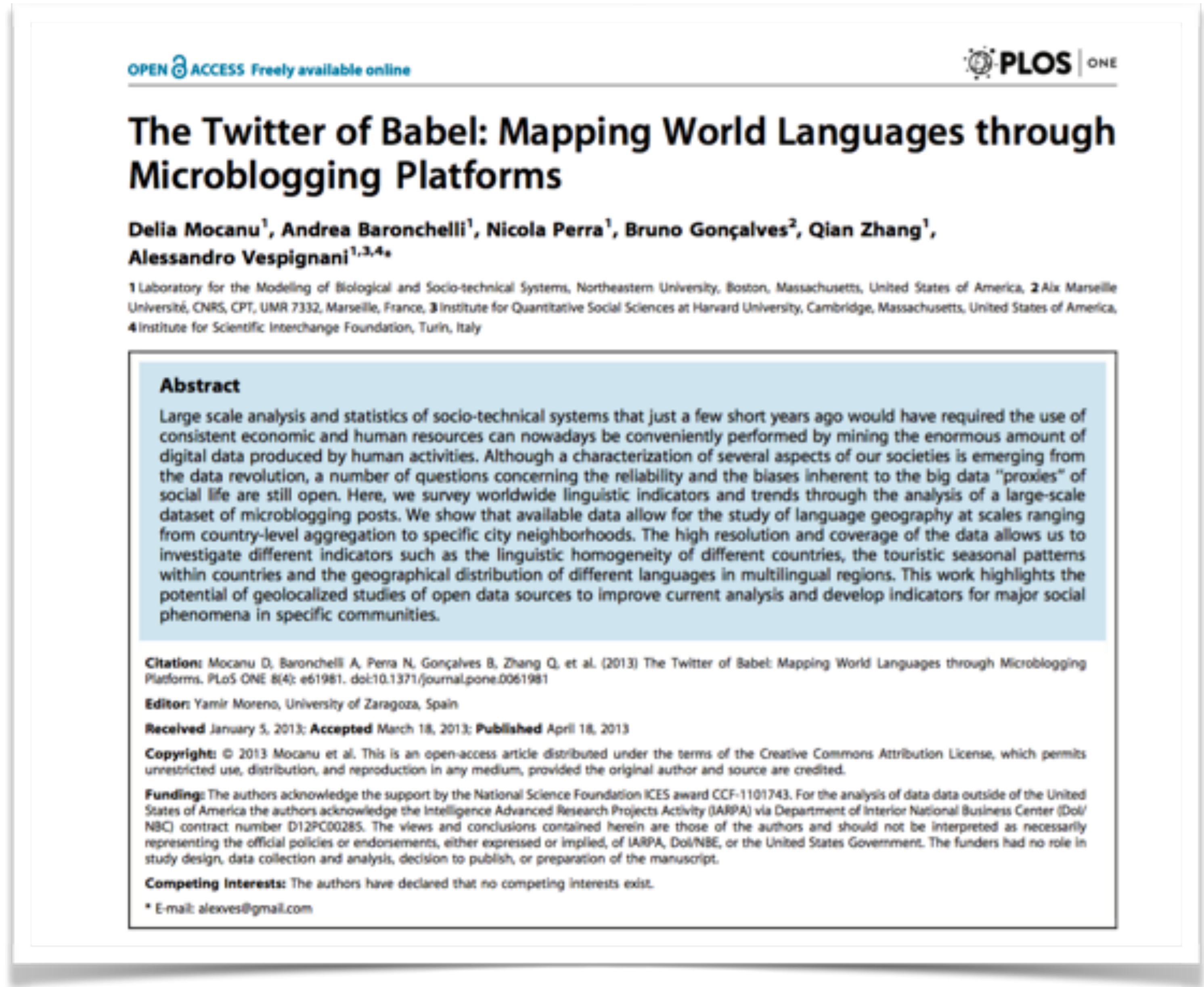
Funding: The authors acknowledge the support by the National Science Foundation ICES award CCF-1101743. For the analysis of data data outside of the United States of America the authors acknowledge the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: alexves@gmail.com

Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981.

LANGUAGE USE



OPEN ACCESS Freely available online

PLOS ONE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

Delia Mocanu[1], Andrea Baronchelli[1], Nicola Perra[1], Bruno Gonçalves[2], Qian Zhang[1], Alessandro Vespignani[1,3,4]*

1 Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, Massachusetts, United States of America, 2 Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France, 3 Institute for Quantitative Social Sciences at Harvard University, Cambridge, Massachusetts, United States of America, 4 Institute for Scientific Interchange Foundation, Turin, Italy

Abstract

Large scale analysis and statistics of socio-technical systems that just a few short years ago would have required the use of consistent economic and human resources can nowadays be conveniently performed by mining the enormous amount of digital data produced by human activities. Although a characterization of several aspects of our societies is emerging from the data revolution, a number of questions concerning the reliability and the biases inherent to the big data "proxies" of social life are still open. Here, we survey worldwide linguistic indicators and trends through the analysis of a large-scale dataset of microblogging posts. We show that available data allow for the study of language geography at scales ranging from country-level aggregation to specific city neighborhoods. The high resolution and coverage of the data allows us to investigate different indicators such as the linguistic homogeneity of different countries, the touristic seasonal patterns within countries and the geographical distribution of different languages in multilingual regions. This work highlights the potential of geolocalized studies of open data sources to improve current analysis and develop indicators for major social phenomena in specific communities.

Citation: Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981

Editor: Yamir Moreno, University of Zaragoza, Spain

Received January 5, 2013; **Accepted** March 18, 2013; **Published** April 18, 2013

Funding: The authors acknowledge the support by the National Science Foundation ICES award CCF-1101743. For the analysis of data data outside of the United States of America the authors acknowledge the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

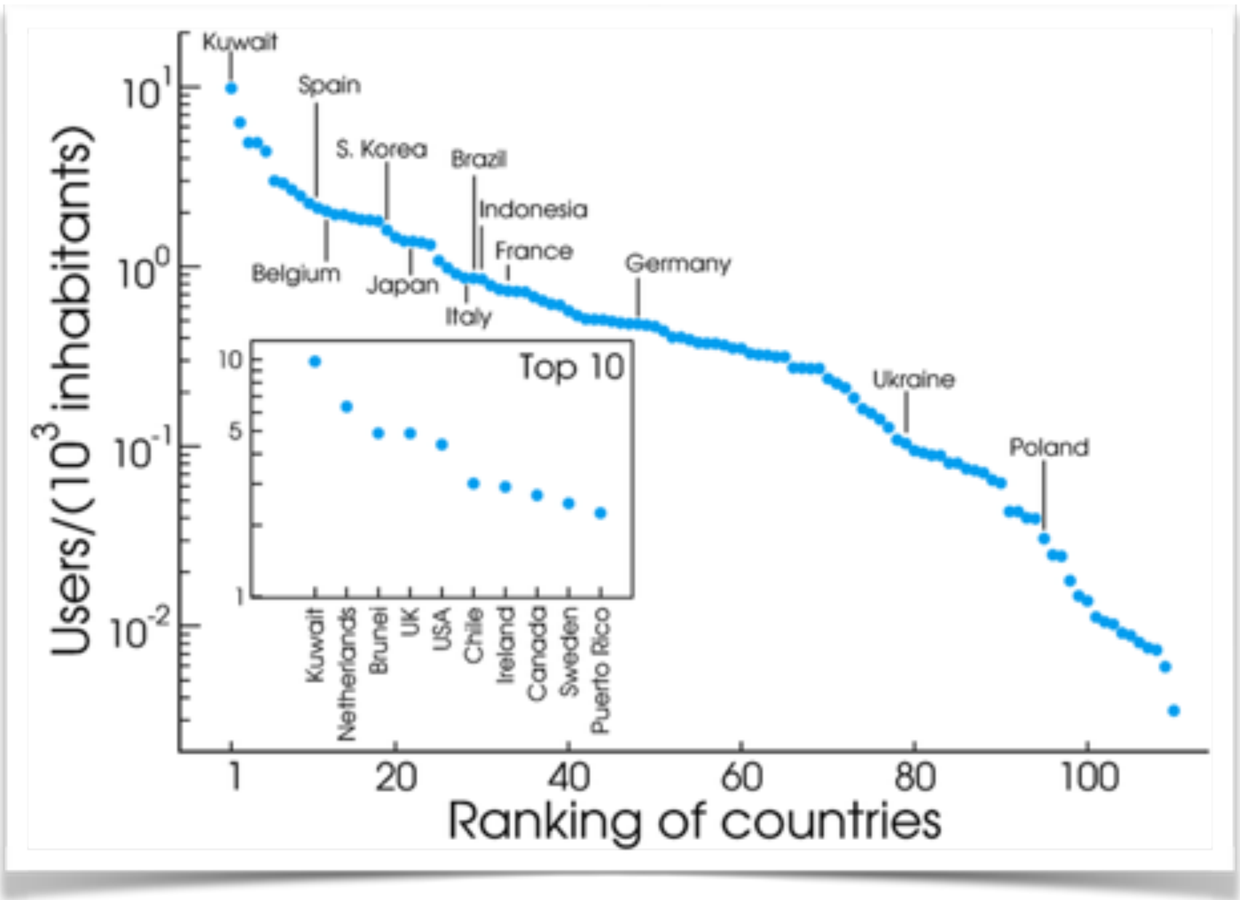
* E-mail: alexves@gmail.com

IS TWITTER REPRESENTATIVE OF OUR SOCIETY?

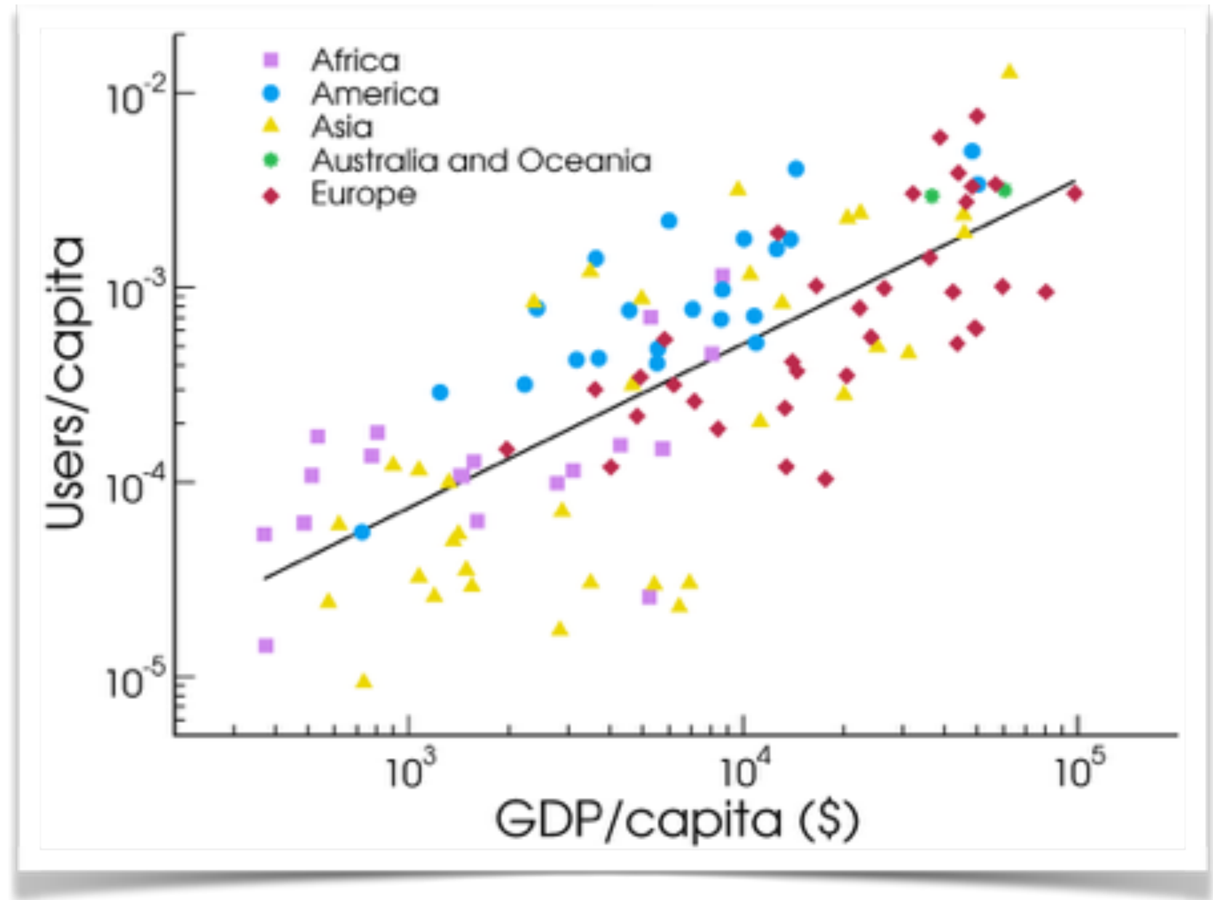
Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981.

LANGUAGE USE

REPRESENTATIVENESS BY COUNTRY



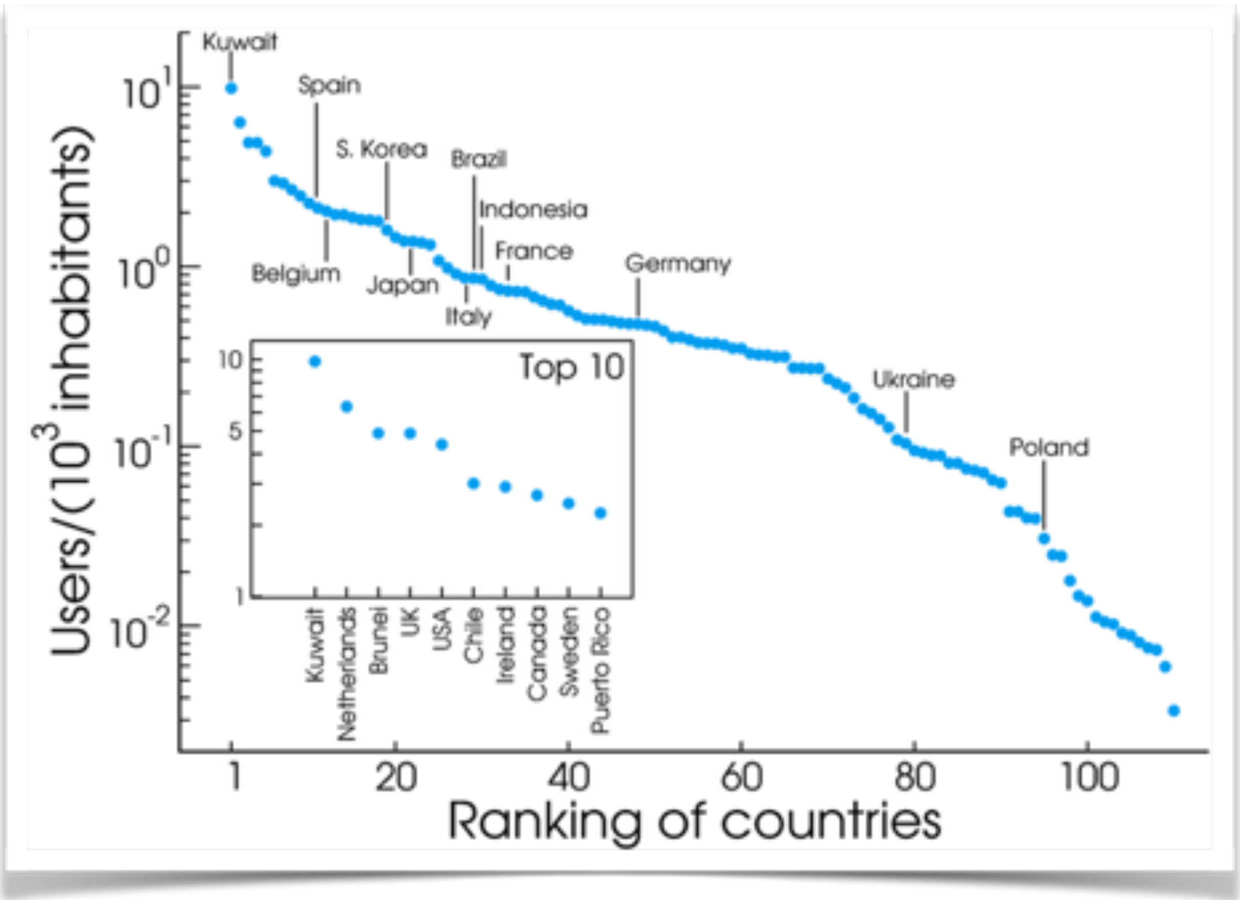
AND BY GDP



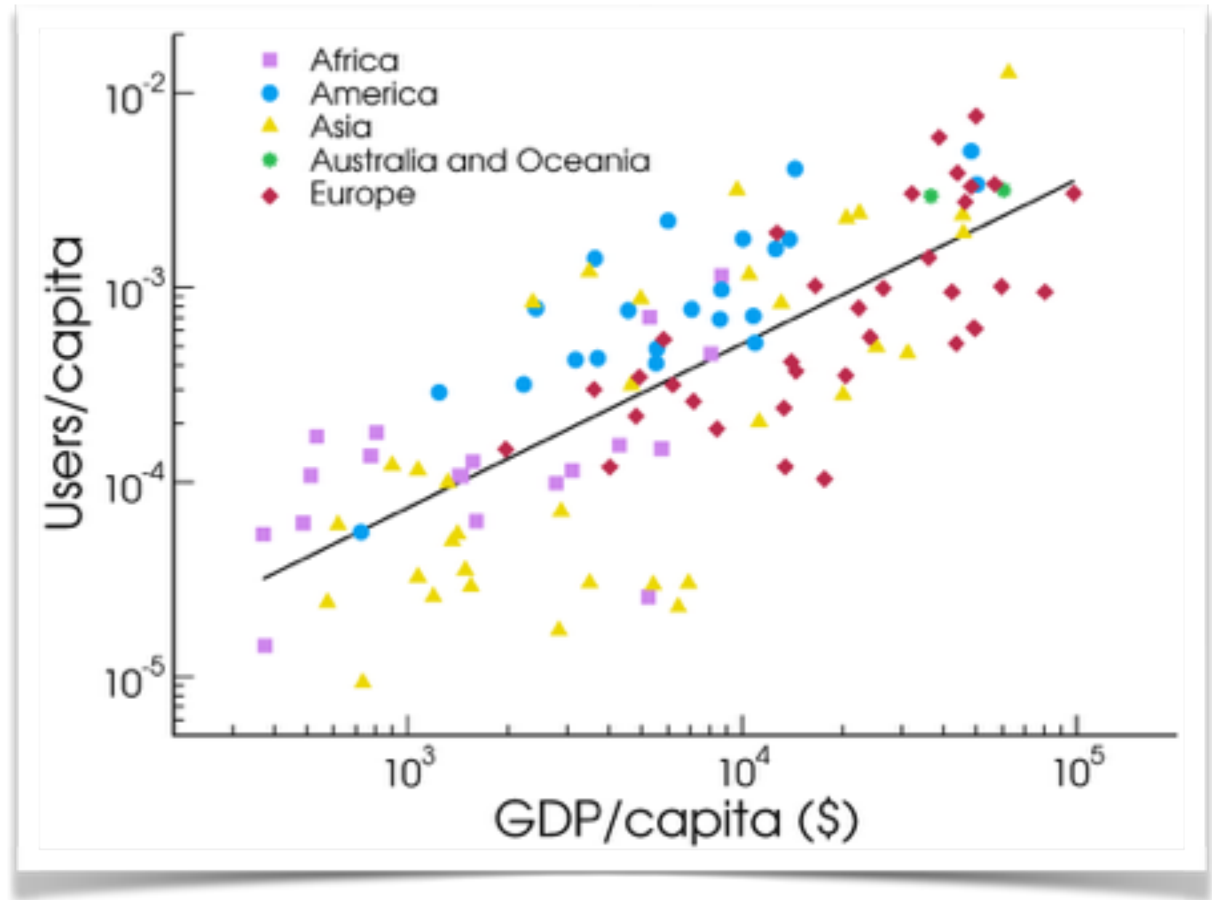
@gomezgardenes

LANGUAGE USE

REPRESENTATIVENESS BY COUNTRY



AND BY GDP

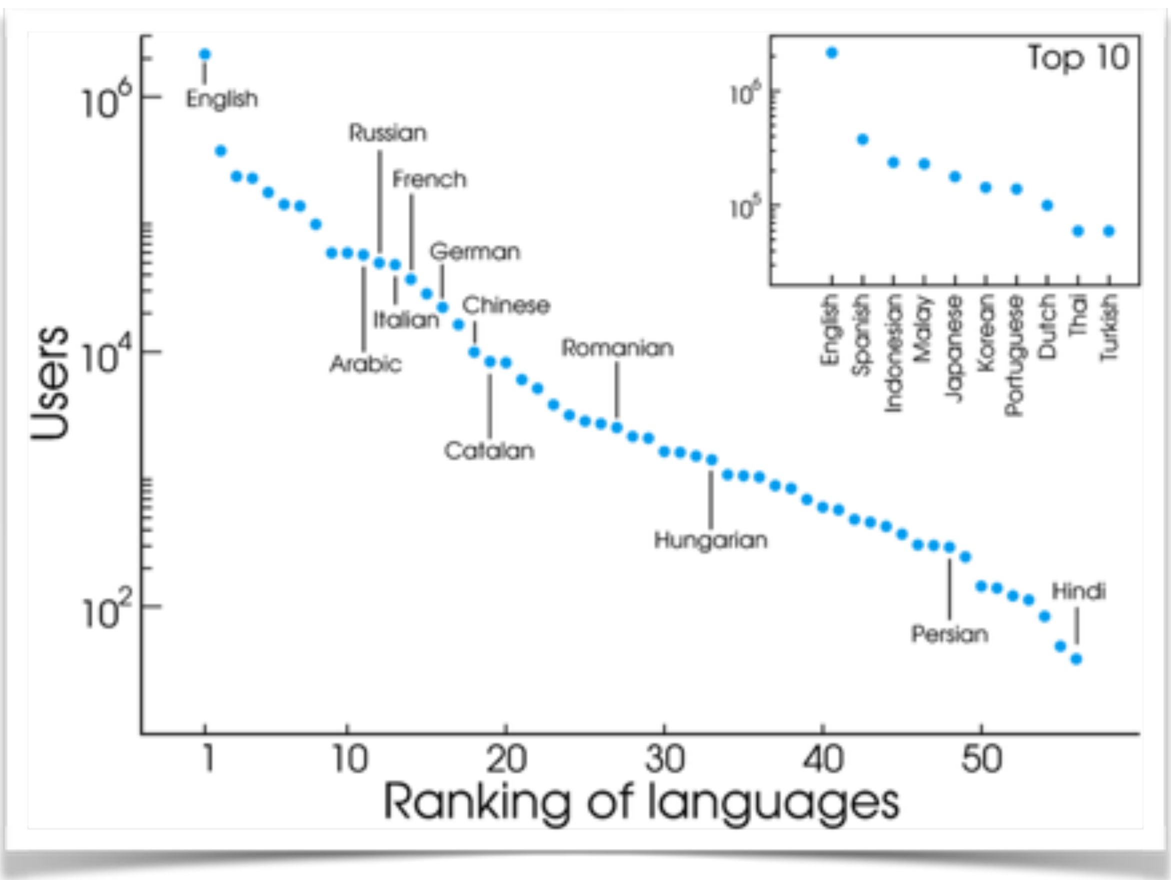


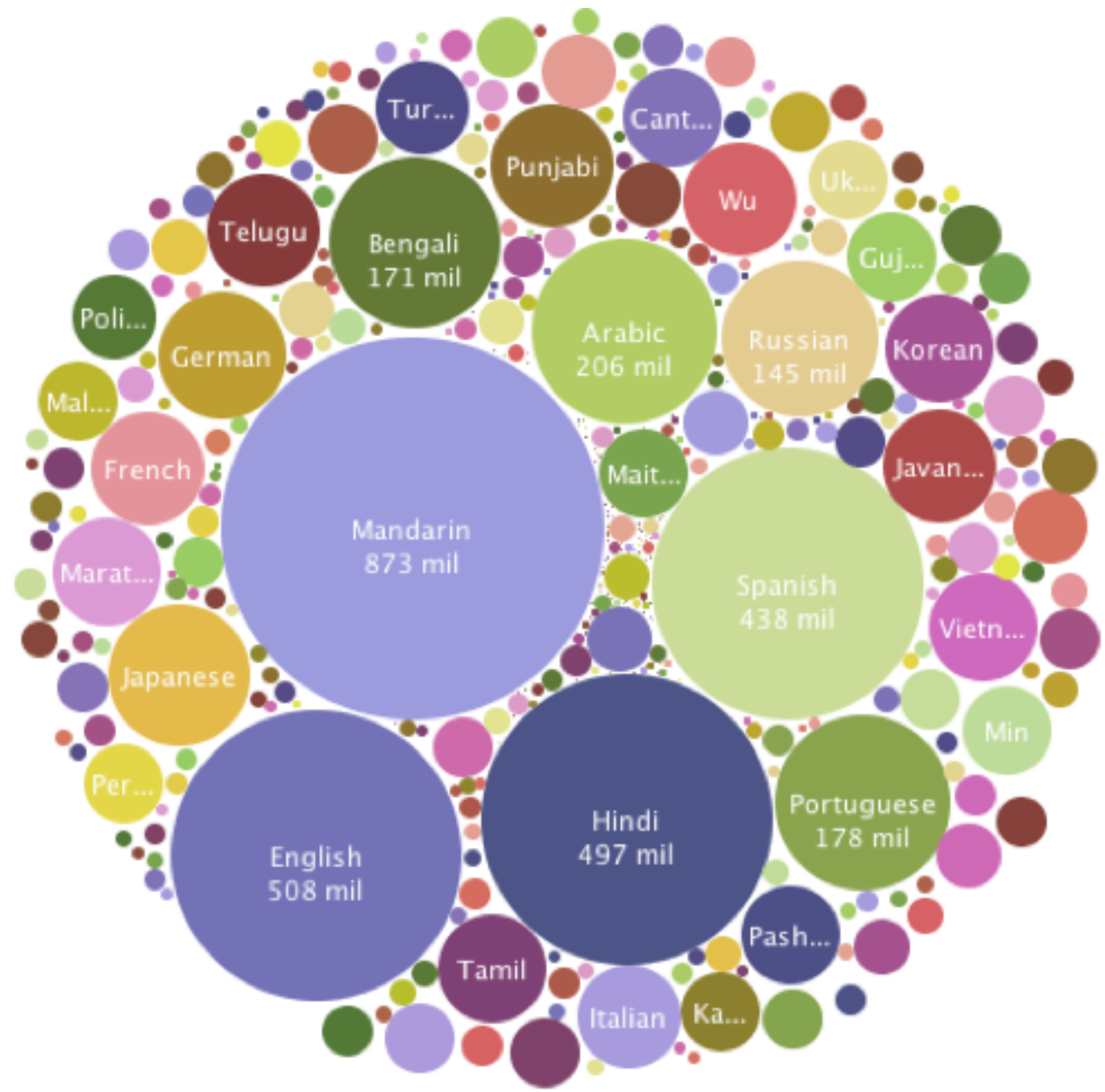
It seems a very good sample for a sociology study
(especially in Kuwait…)

@gomezgardenes

LANGUAGE USE

LANGUAGES

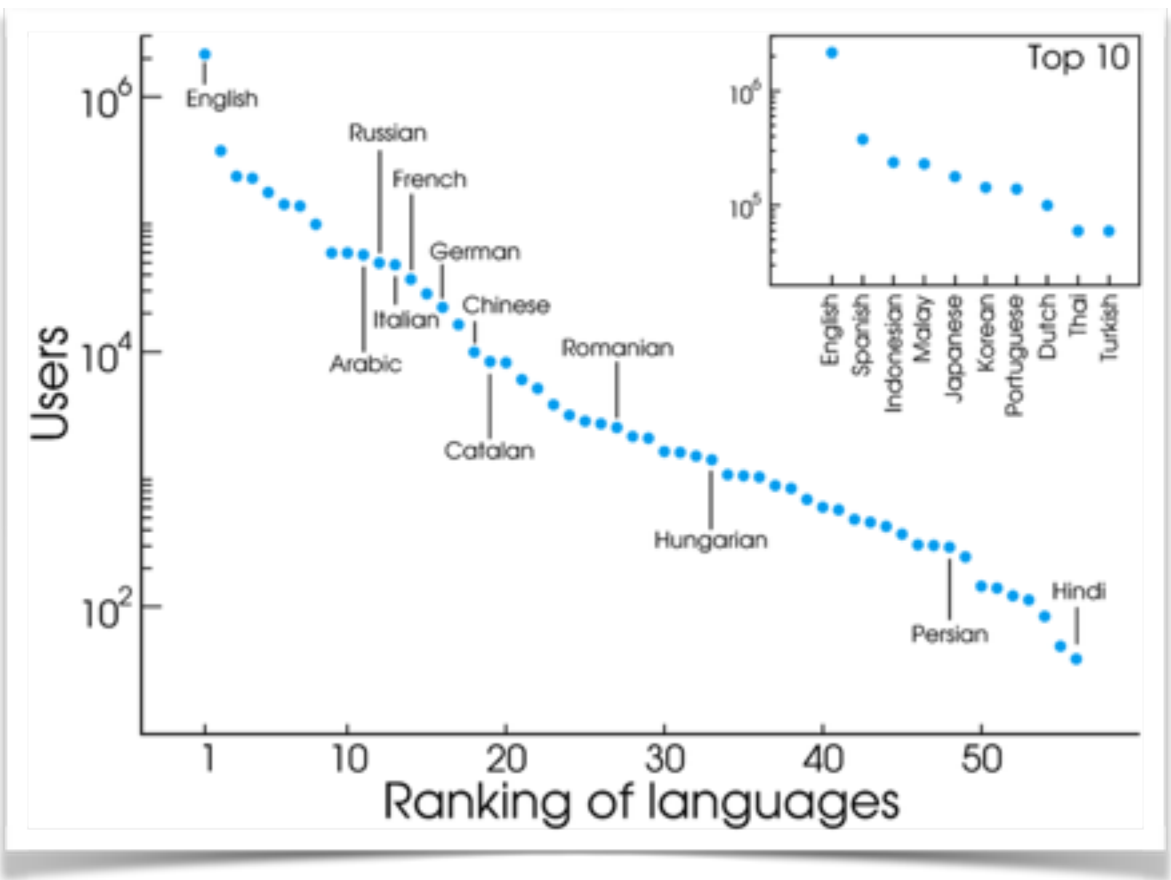


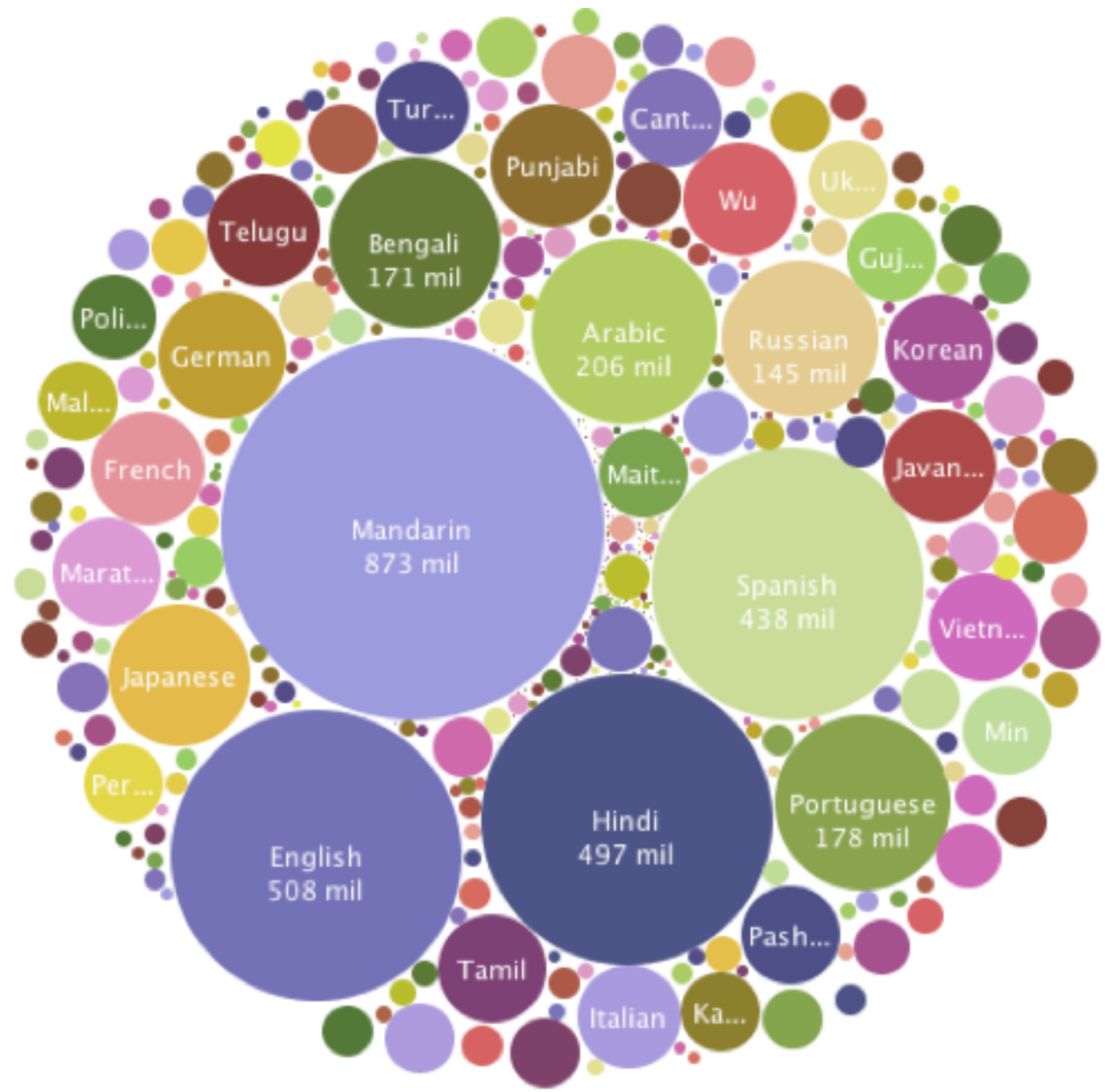


@gomezgardenes

LANGUAGE USE

LANGUAGES



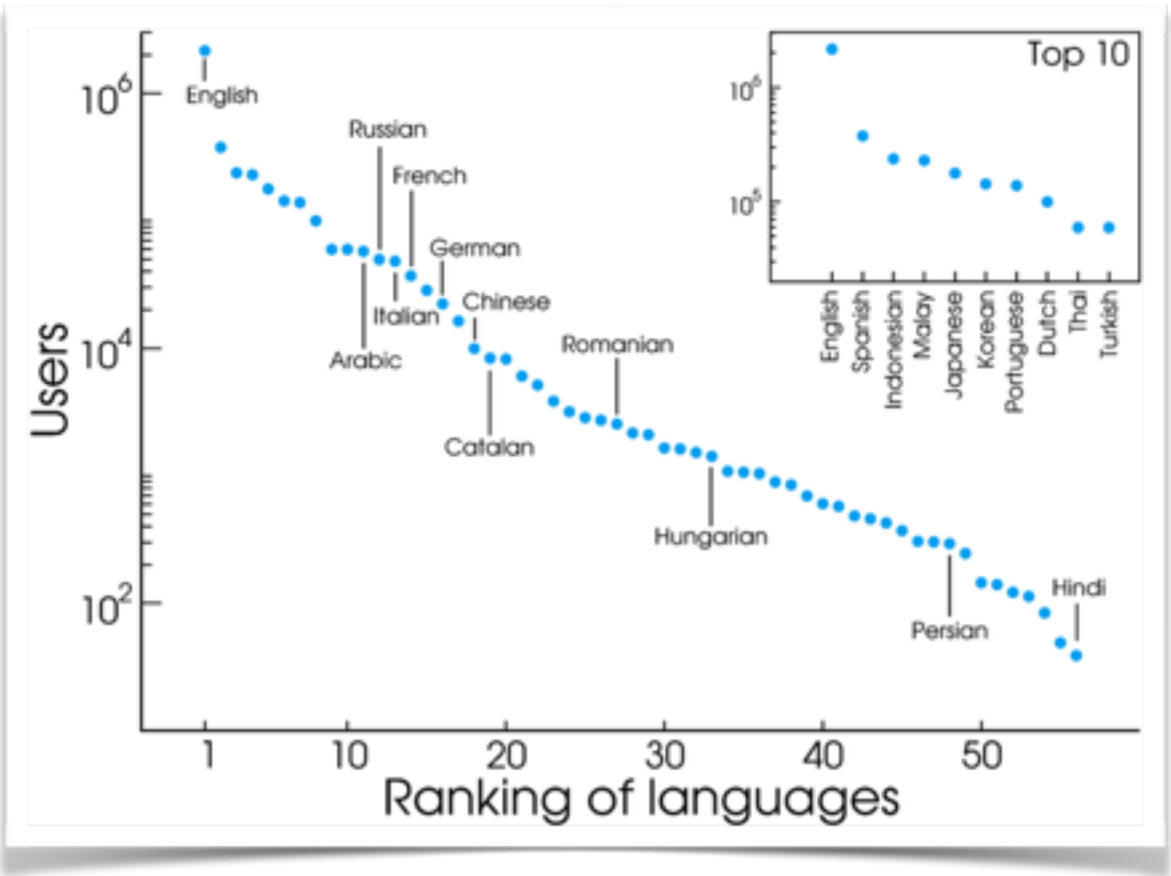


English overrepresented

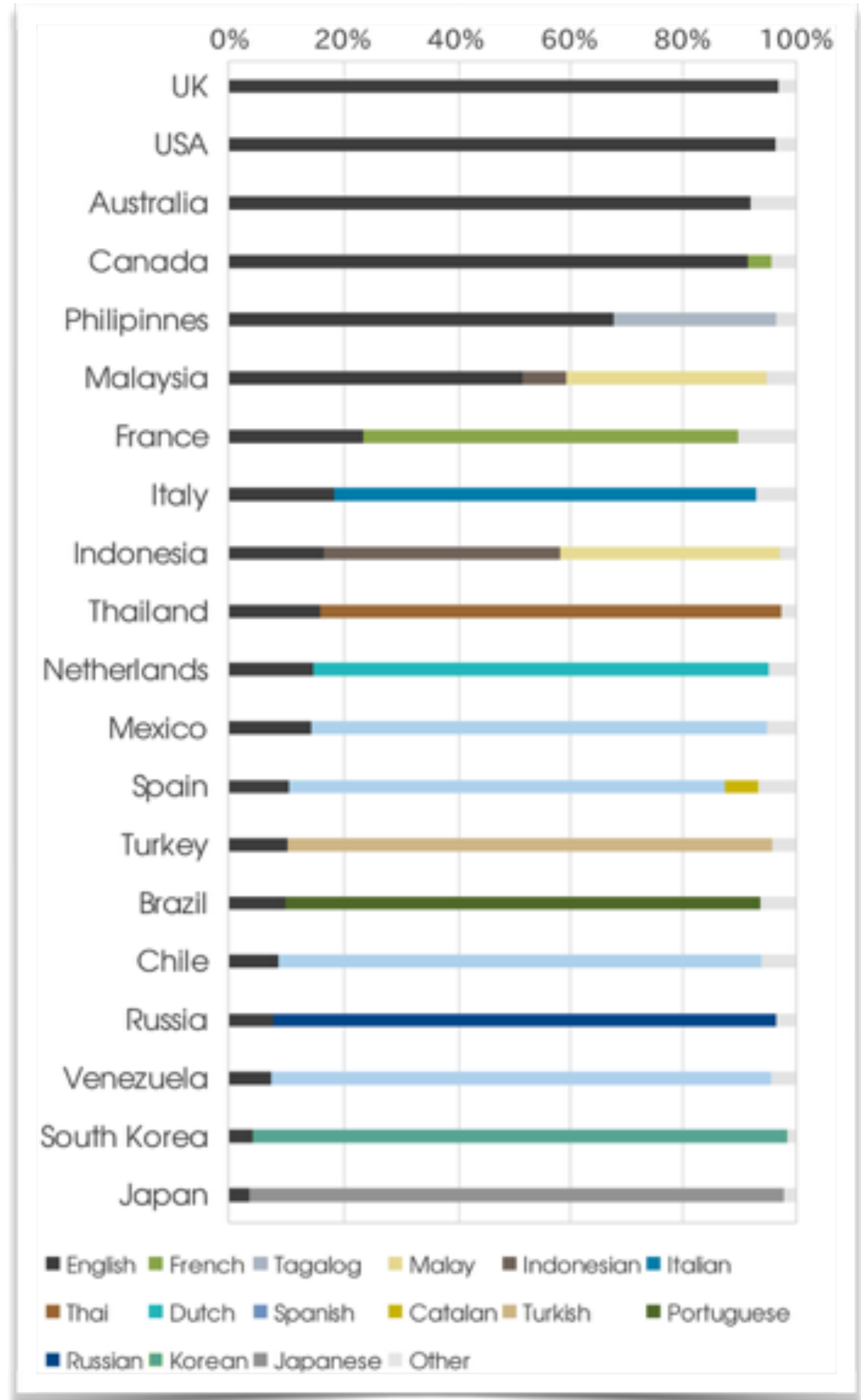
@gomezgardenes

LANGUAGE USE

LANGUAGES



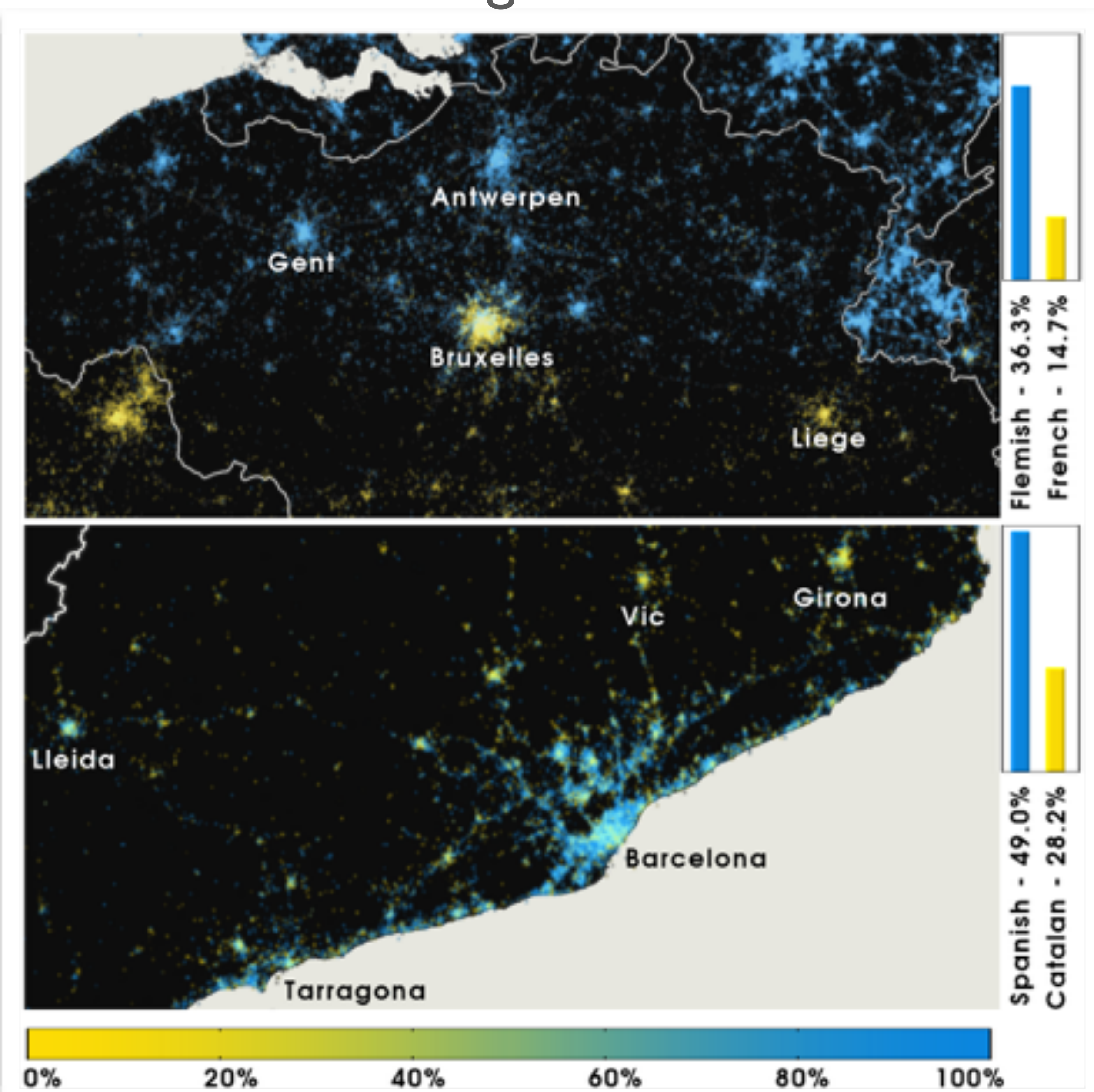
English overrepresented



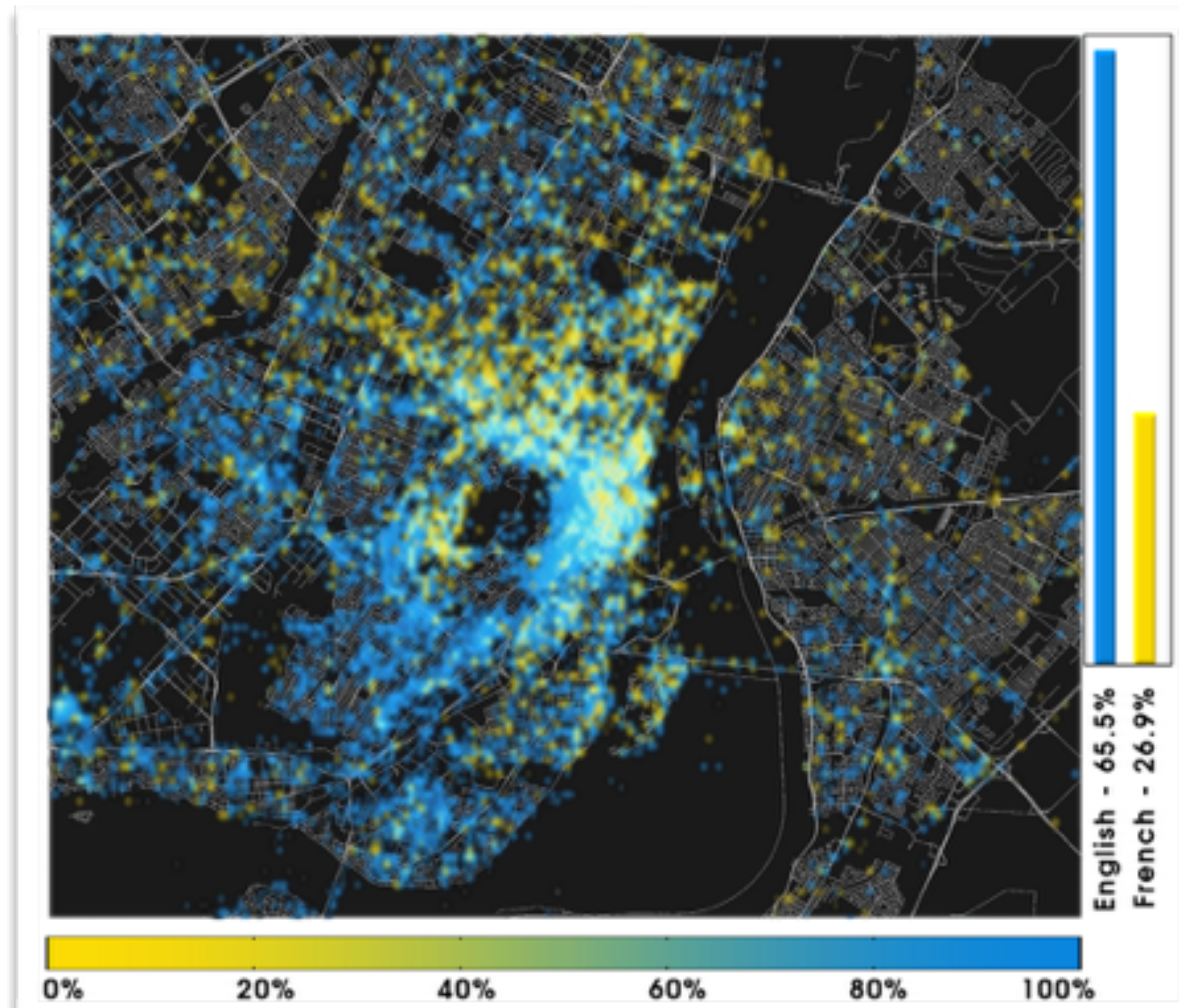
@gomezgardenes

LANGUAGES POLARIZATION

Belgium



Catalonia, ES



Montreal, CA

@gomezgardenes

DIGITAL EPIDEMIOLOGY

nature Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\text{logit}(I(t)) = \alpha\text{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\text{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant. Detecting influenza epidemics using search engine query data Nature 457, 1012-1014 (19 February 2009)

@gomezgardenes

DIGITAL EPIDEMIOLOGY

The initial Google paper stated that the Google Flu Trends predictions were 97% accurate comparing with CDC data.

nature Vol 457|19 February 2009|**doi:10.1038/nature07634**

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\mathrm{logit}(I(t)) = \alpha\,\mathrm{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\mathrm{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant. Detecting influenza epidemics using search engine query data Nature 457, 1012-1014 (19 February 2009)

@gomezgardenes

DIGITAL EPIDEMIOLOGY

The initial Google paper stated that the Google Flu Trends predictions were 97% accurate comparing with CDC data.

nature Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\text{logit}(I(t)) = \alpha\,\text{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\text{logit}(p)$ is simply $\ln(p/(1-p))$.

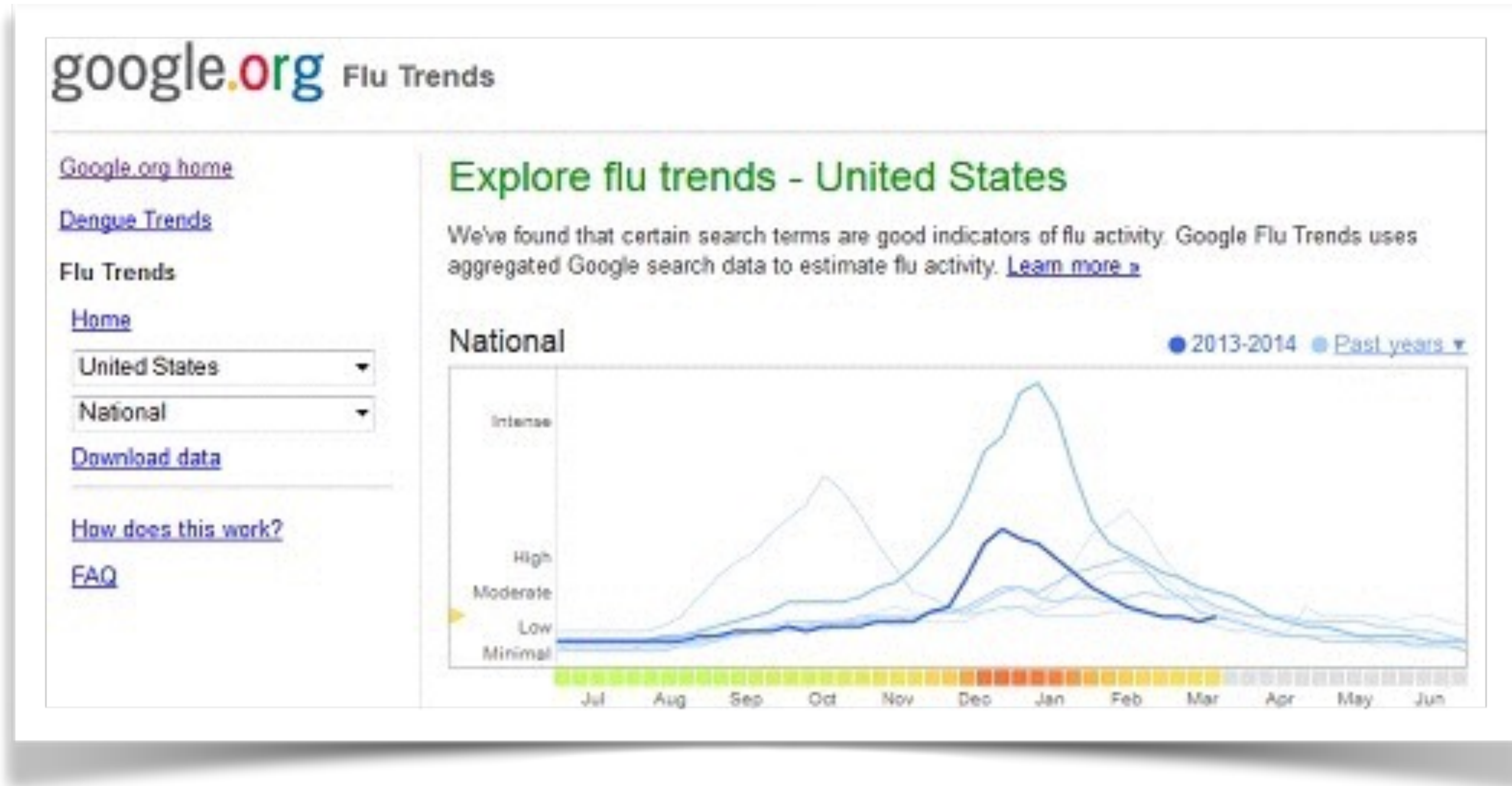
Publicly available historical data from the CDC's US Influenza

Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant. Detecting influenza epidemics using search engine query data Nature 457, 1012-1014 (19 February 2009)

DIGITAL EPIDEMIOLOGY

GOOGLE FLU TRENDS

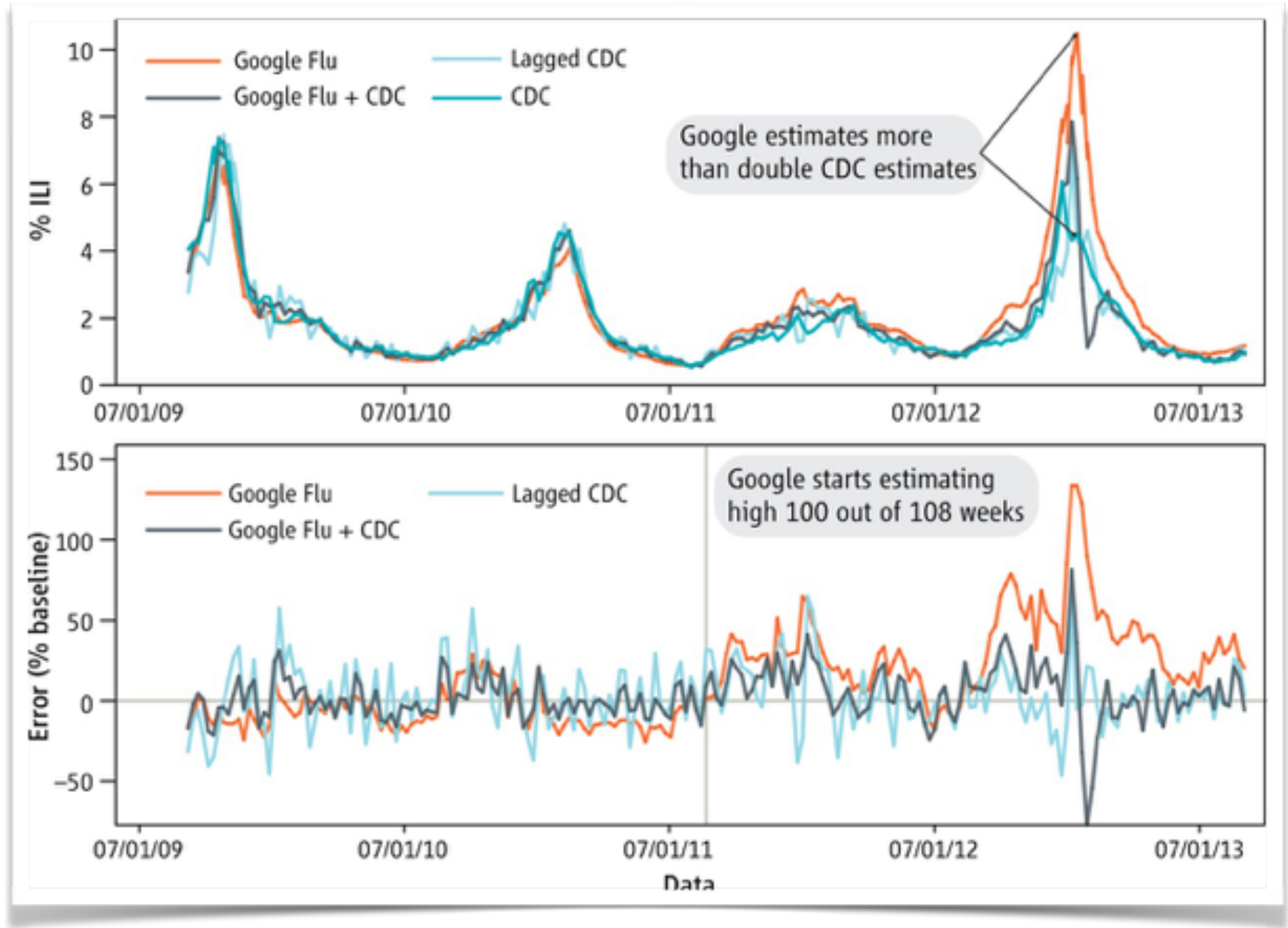
- First launched in 2008 by Google.org to help predict outbreaks of flu.
- More than 25 countries
- First example of "Big Data" for social/health use



The idea behind Google Flu Trends (GFT) is that, by monitoring millions of users' health tracking behaviors online, the large number of Google search queries gathered can be analyzed to reveal if there is the presence of flu-like illness in a population.

The infamous 2012-2013 season

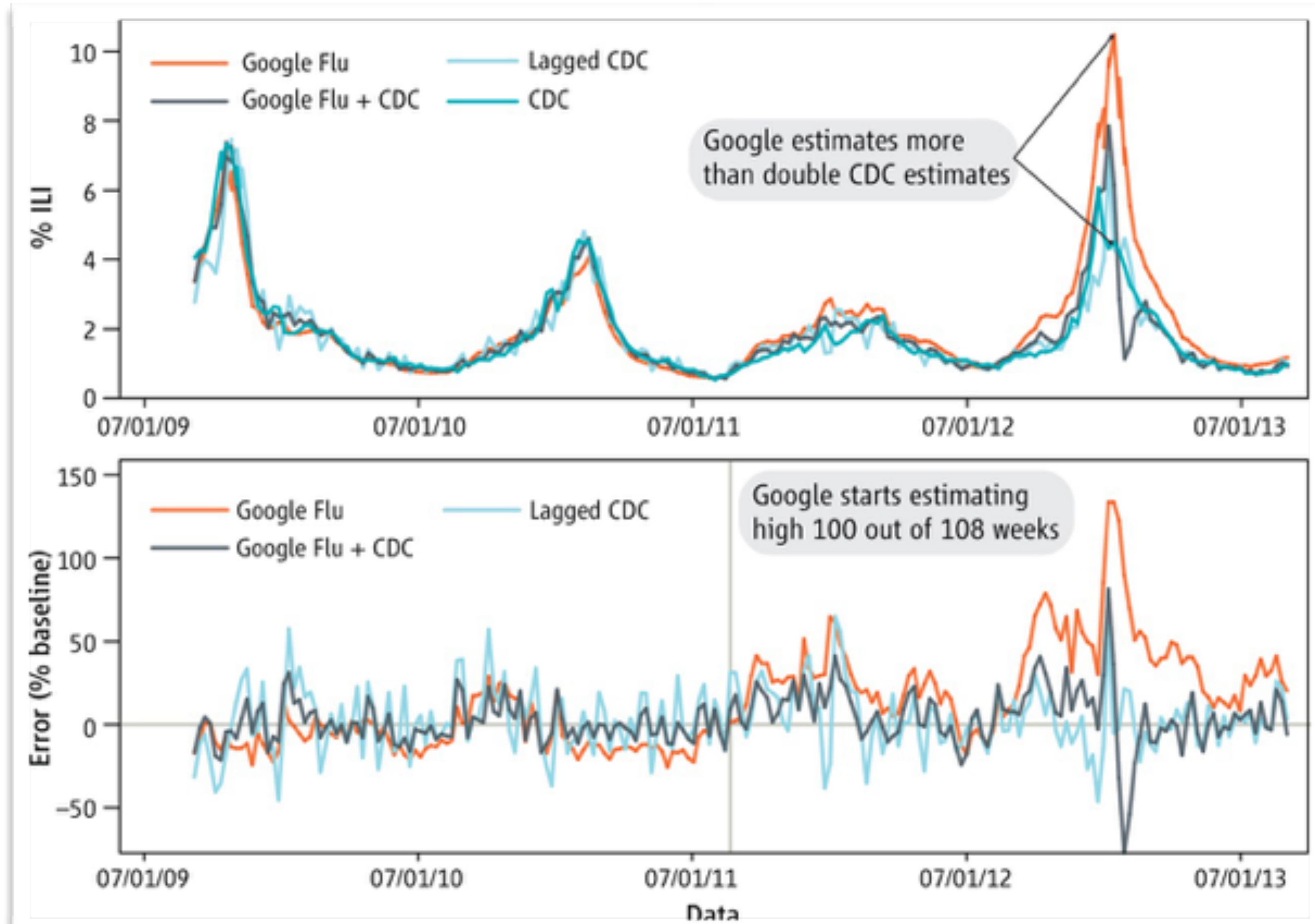
WHAT HAPPENED ?



The infamous 2012-2013 season

WHAT HAPPENED ?

- Early Peak season
- Highly contagious strain (H3N2)
- Huge news coverage



BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

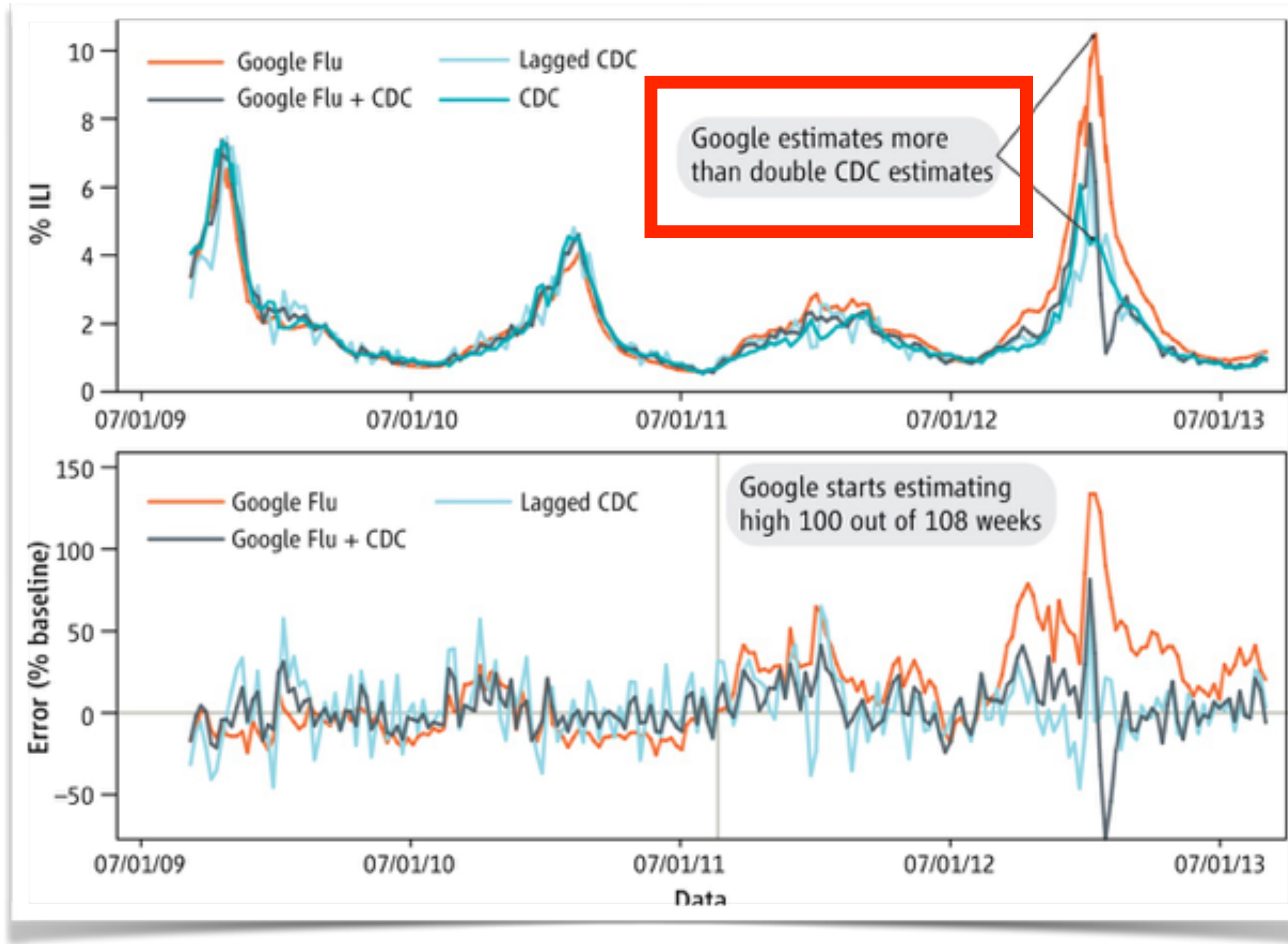
David Lazer,[1,2]* Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

Science 343, 6176, 1203-1205 (2014)

The infamous 2012-2013 season

WHAT HAPPENED ?

- Early Peak season
- Highly contagious strain (H3N2)
- Huge news coverage



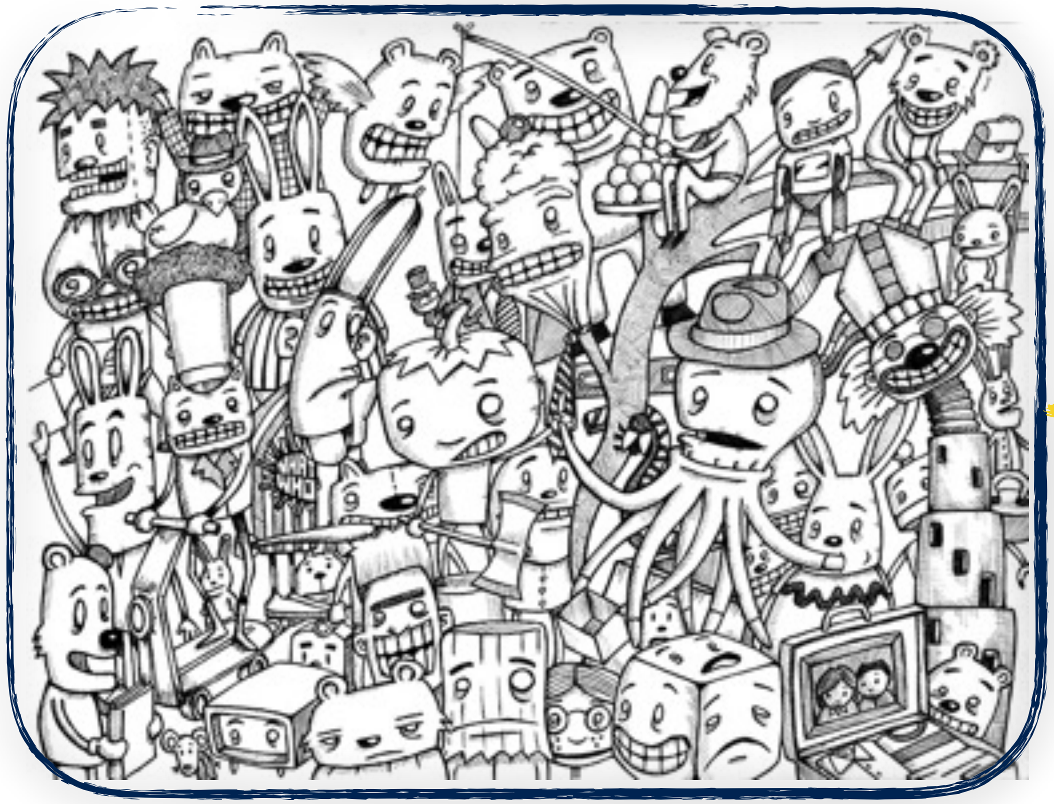
BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2]* Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

Science 343, 6176, 1203-1205 (2014)

@gomezgardenes



+ Realism

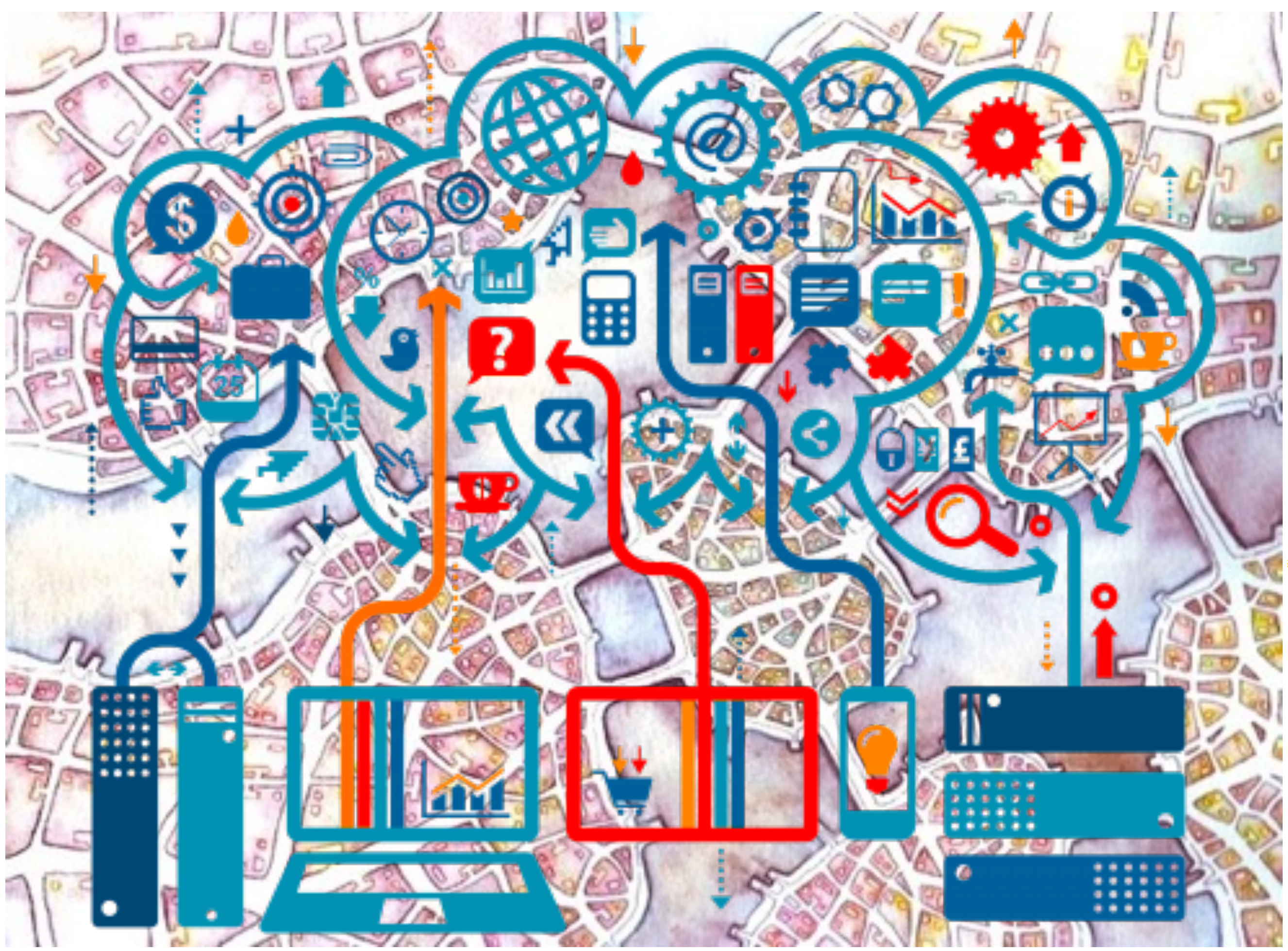
+ Complexity

+ Realism

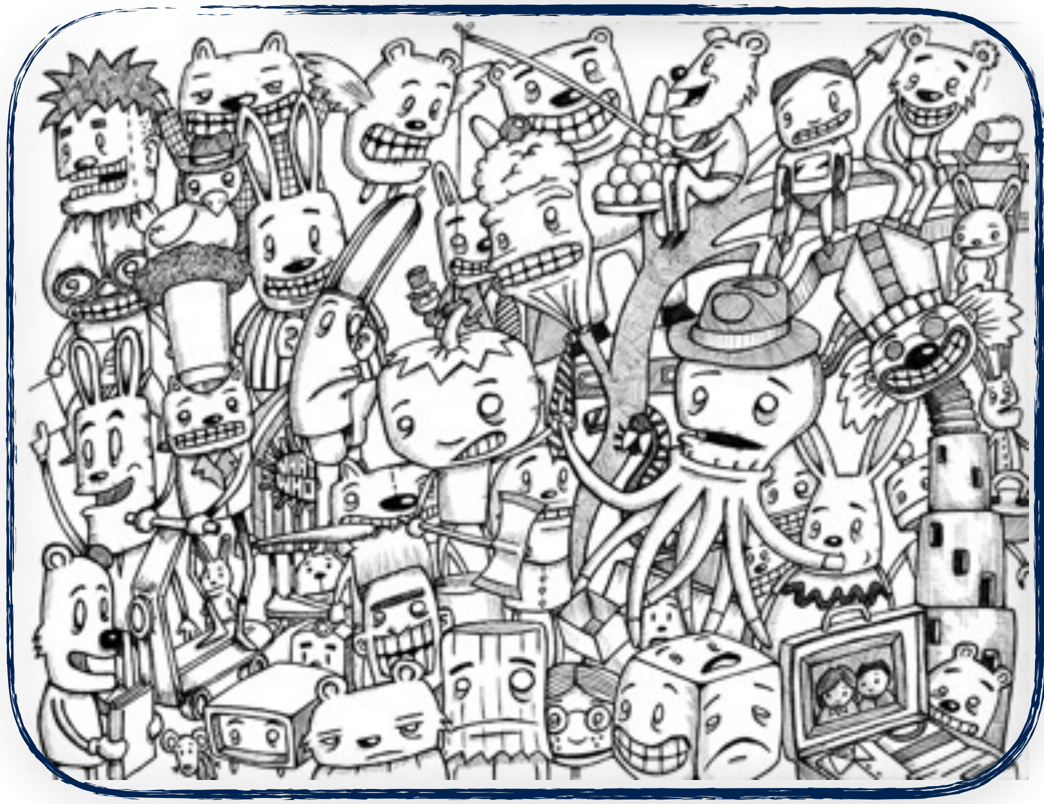
+ Realism

+ Complexity

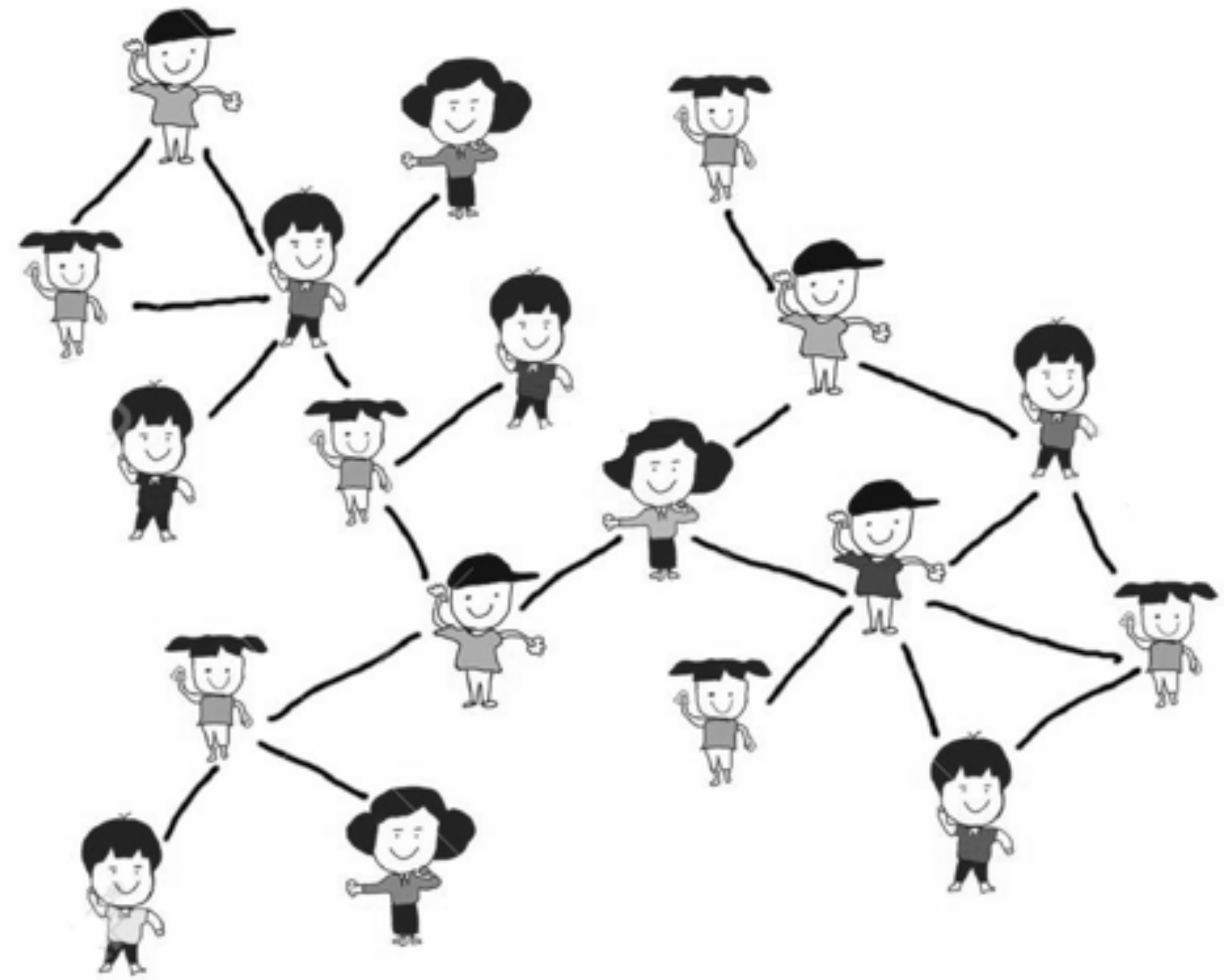
+ Realism

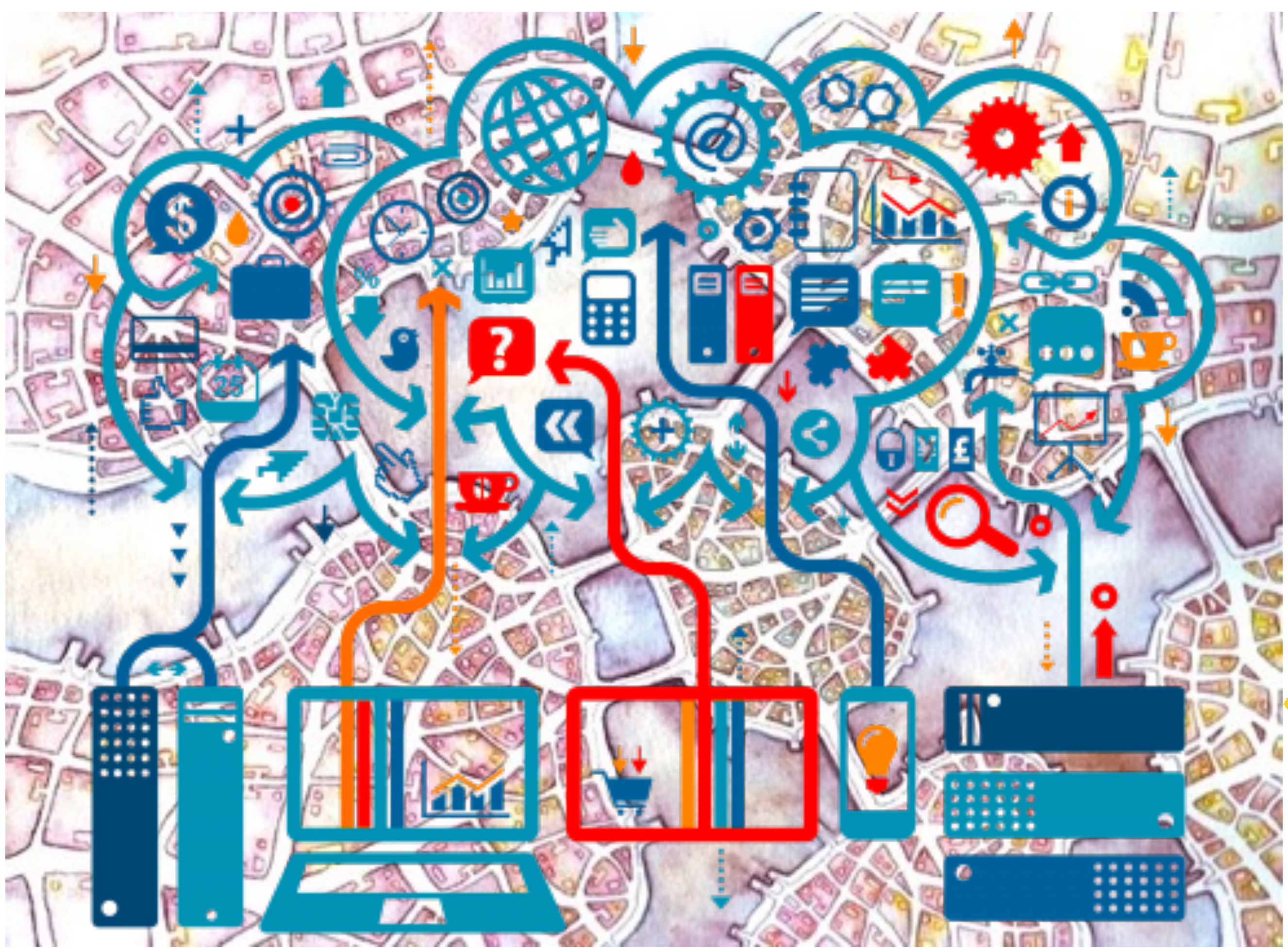


@gomezgardenes



+ Realism
+ Complexity

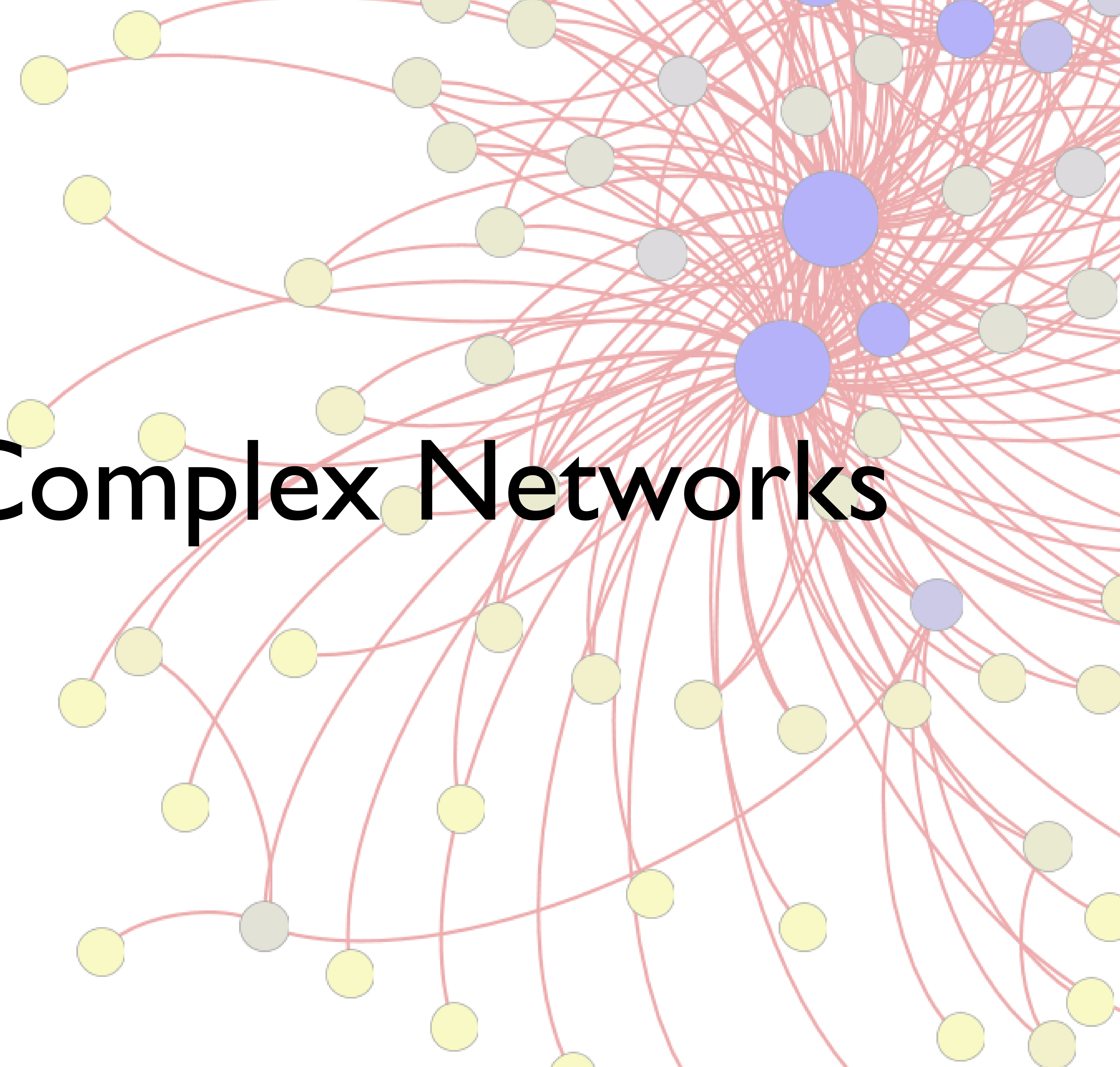




+ Realism
+ Complexity

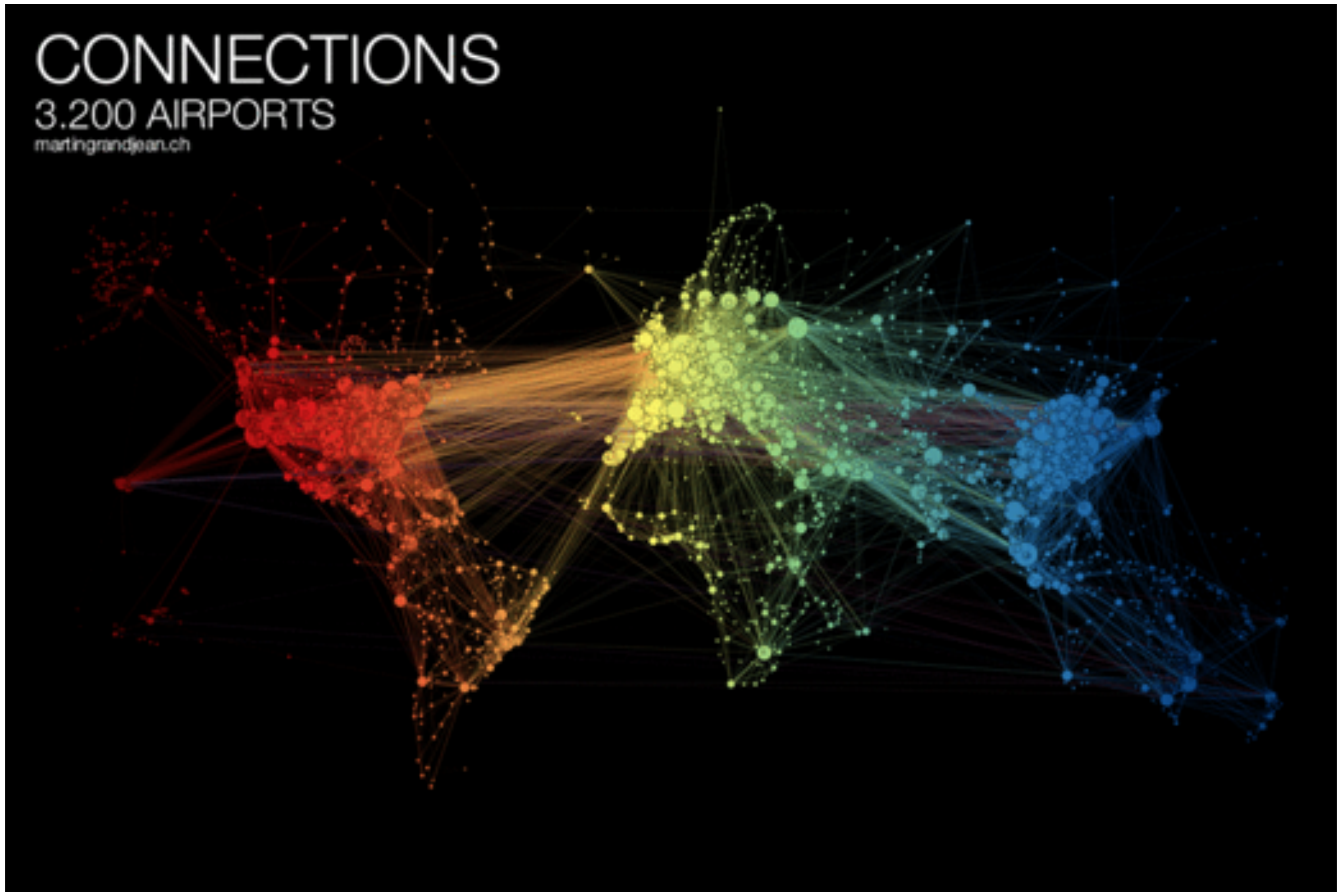
@gomezgardenes

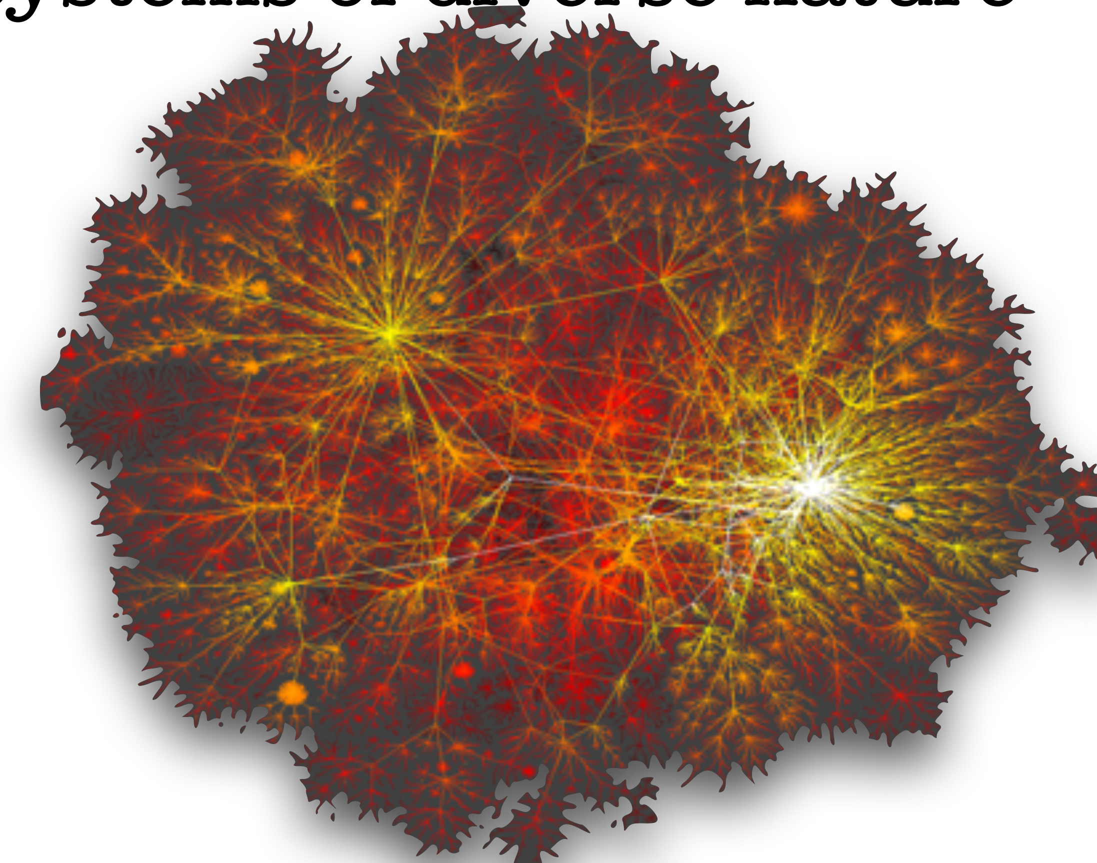
Complex Networks

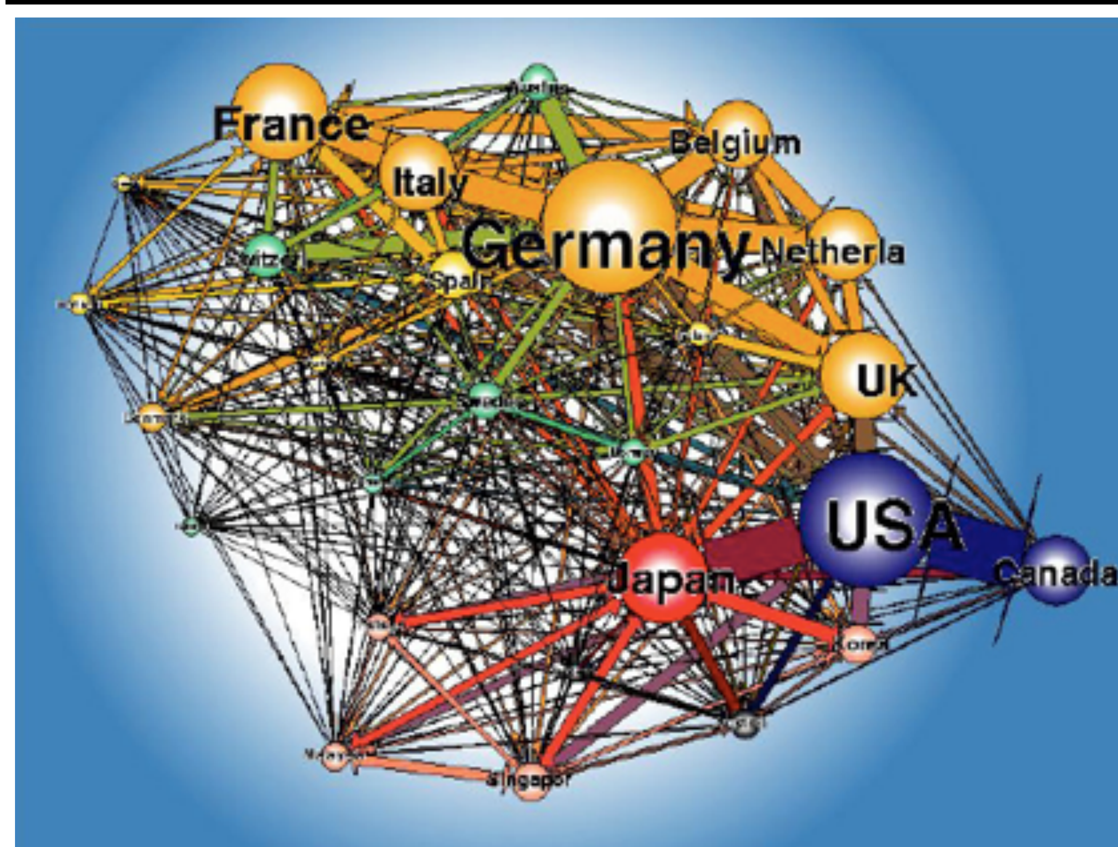


@gomezgardenes

Common language for complex systems of diverse nature

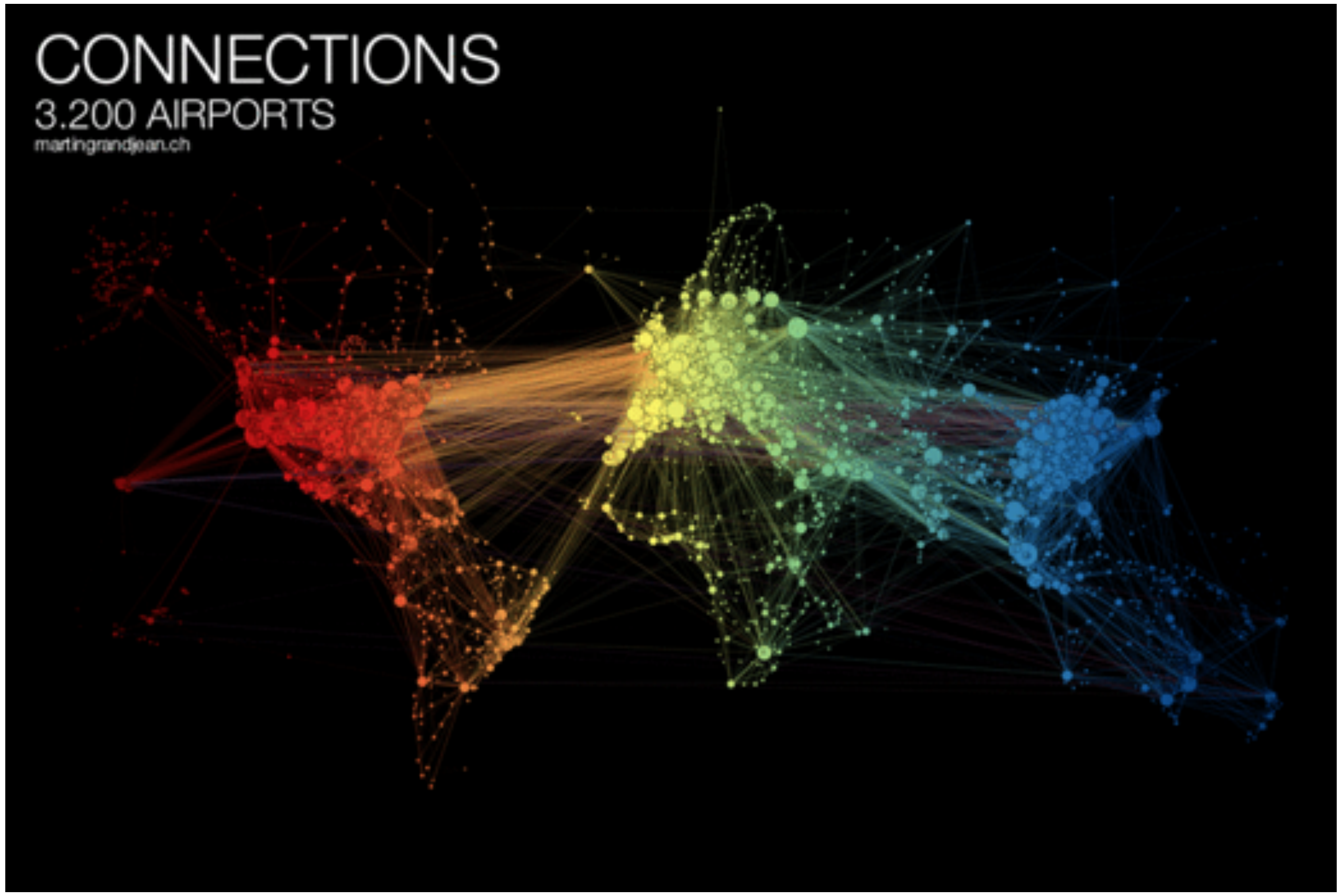


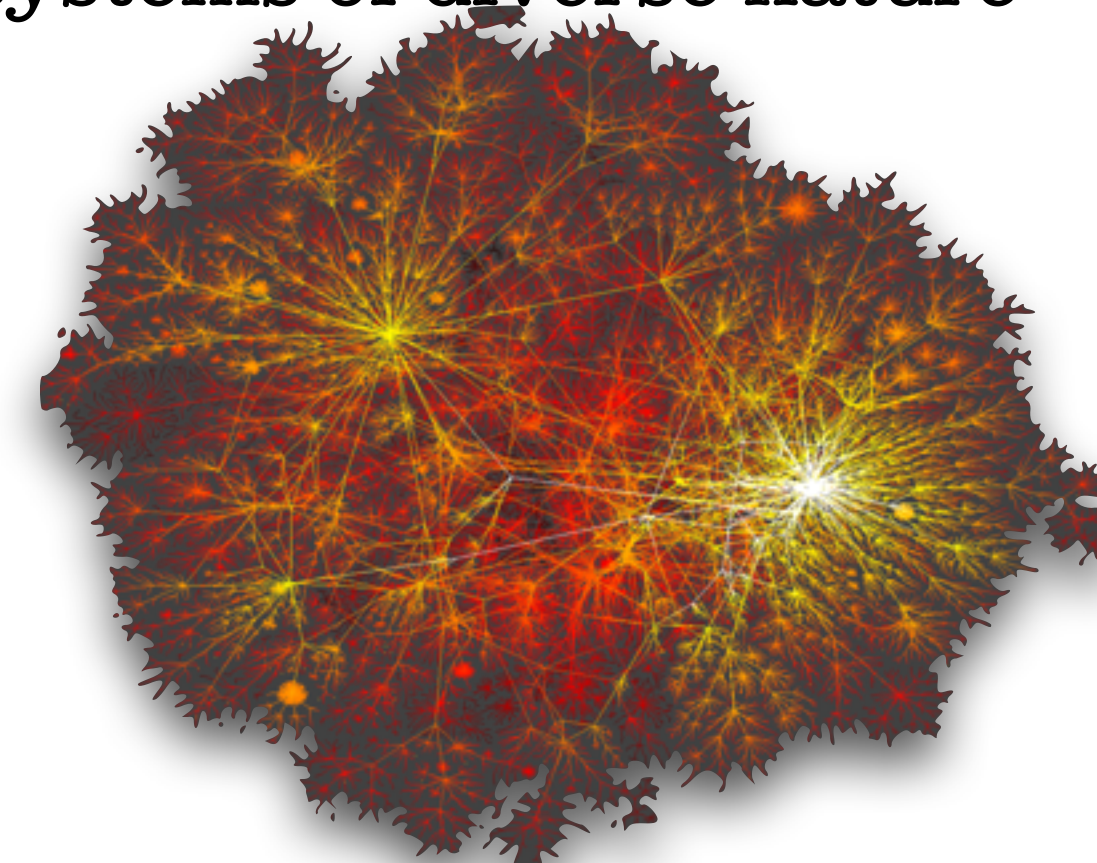


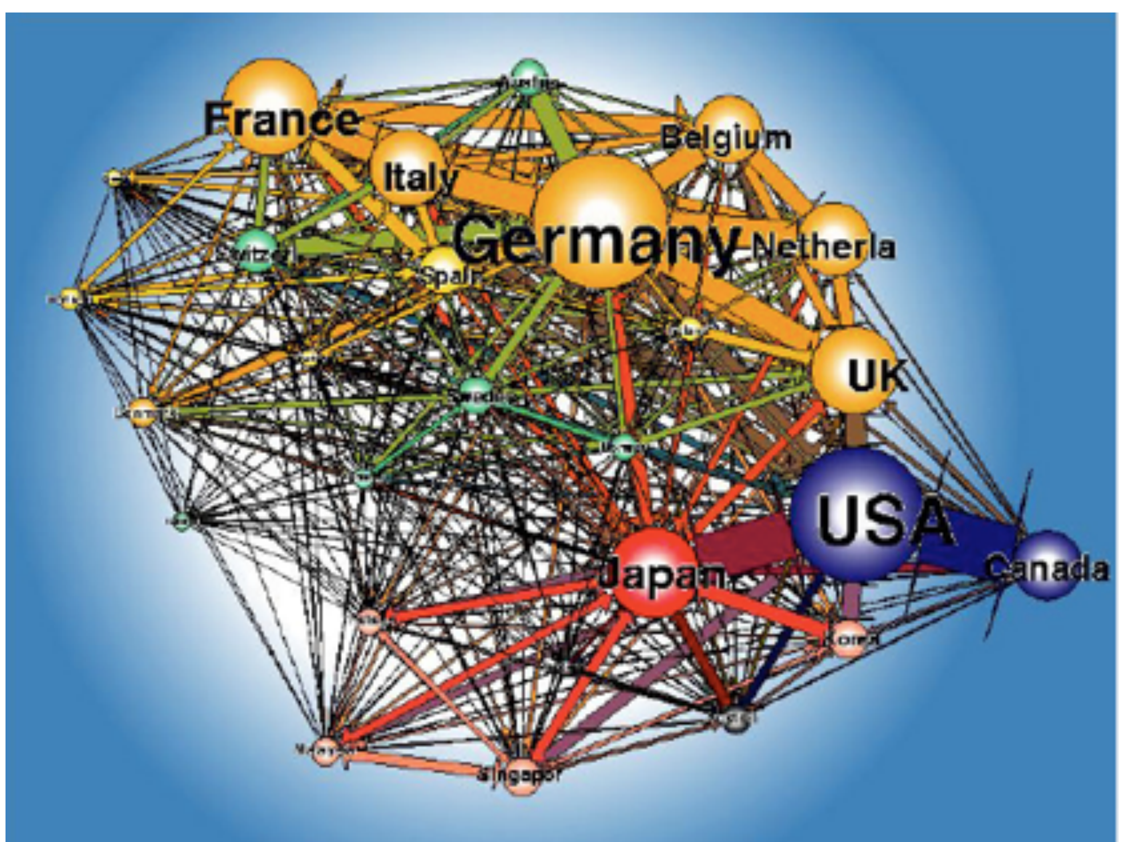


@gomezgardenes

Common language for complex systems of diverse nature

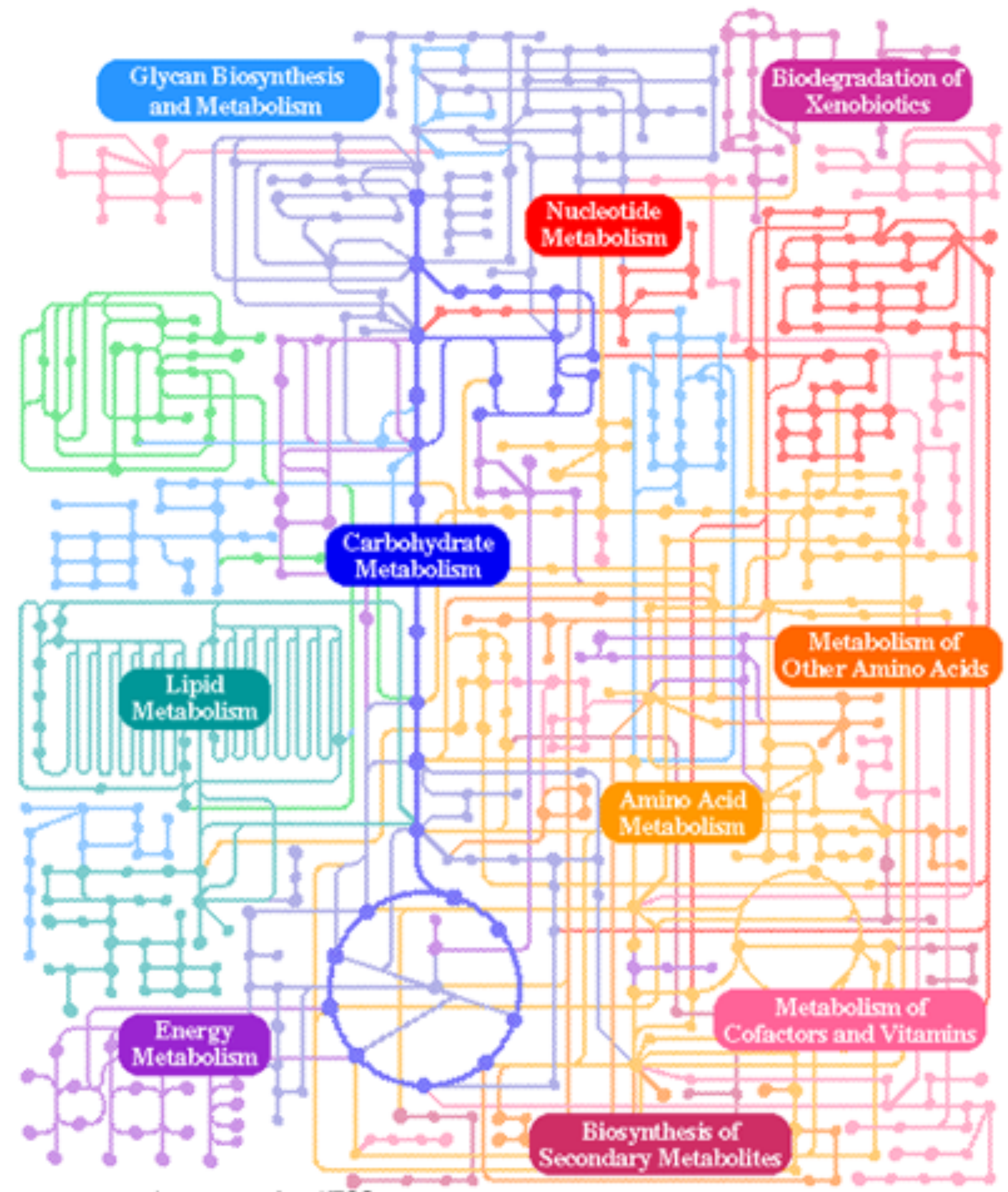




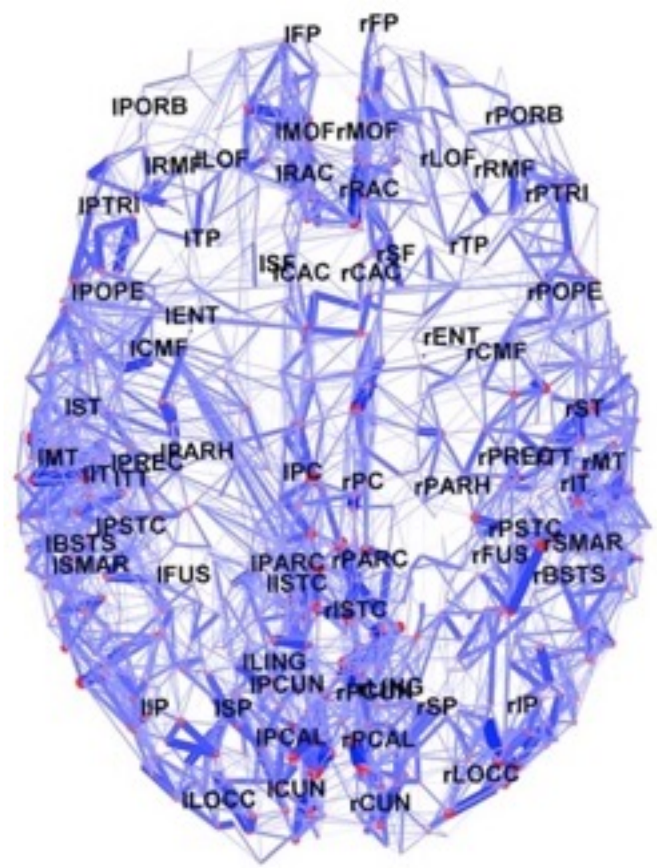


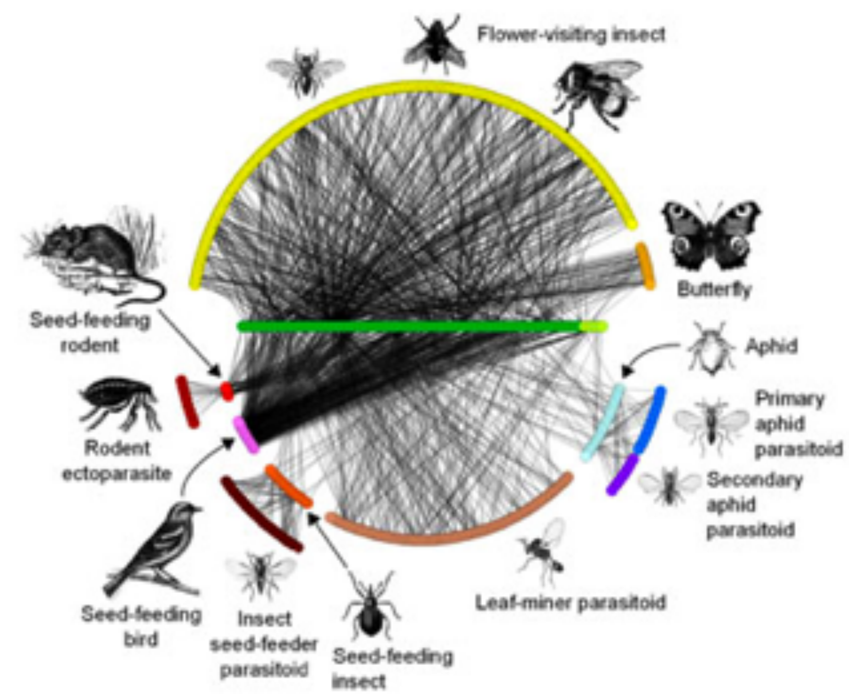
@gomezgardenes

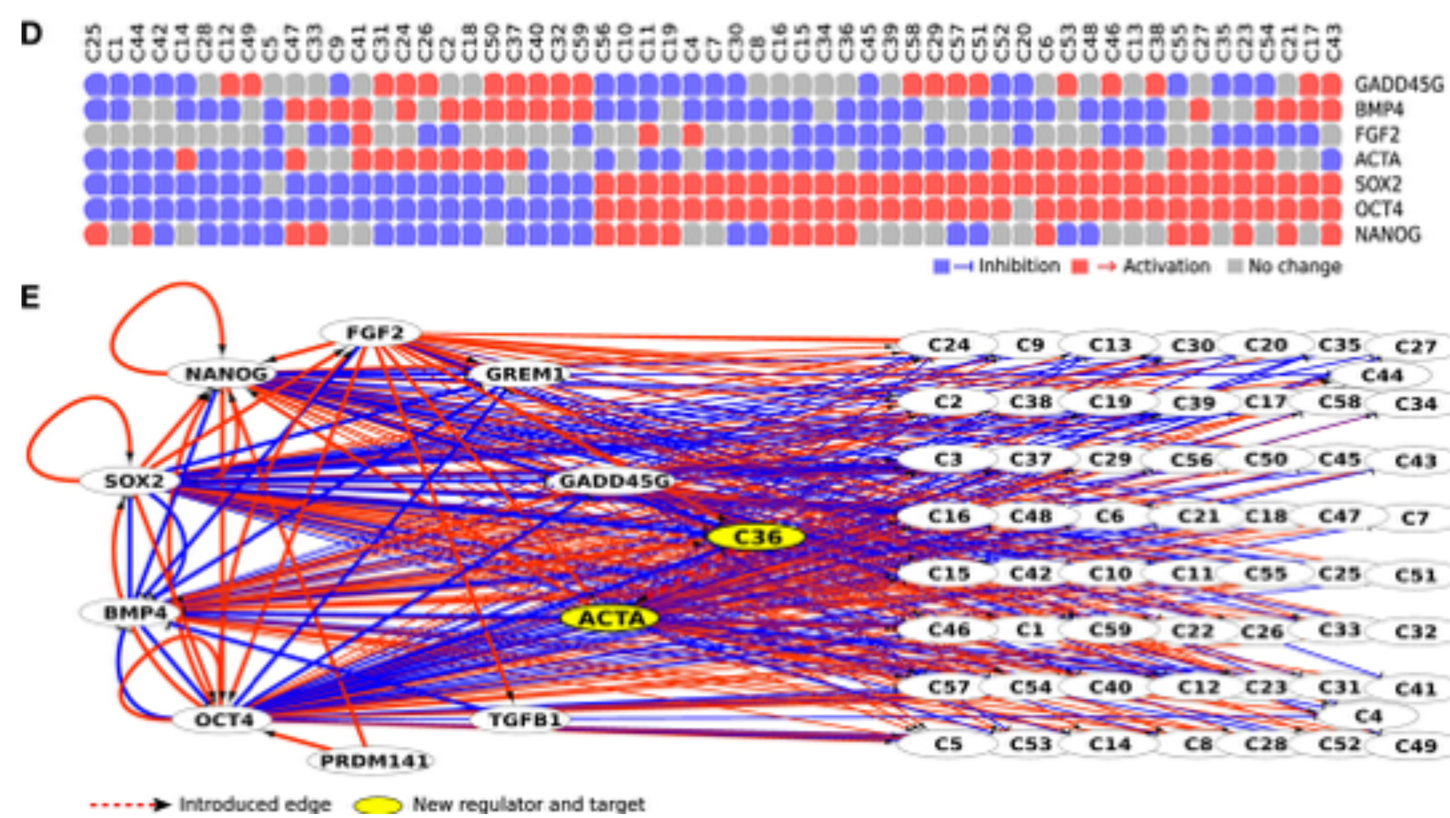
Common language for complex systems of diverse nature



01100 5/31/04 Image source from KEGG

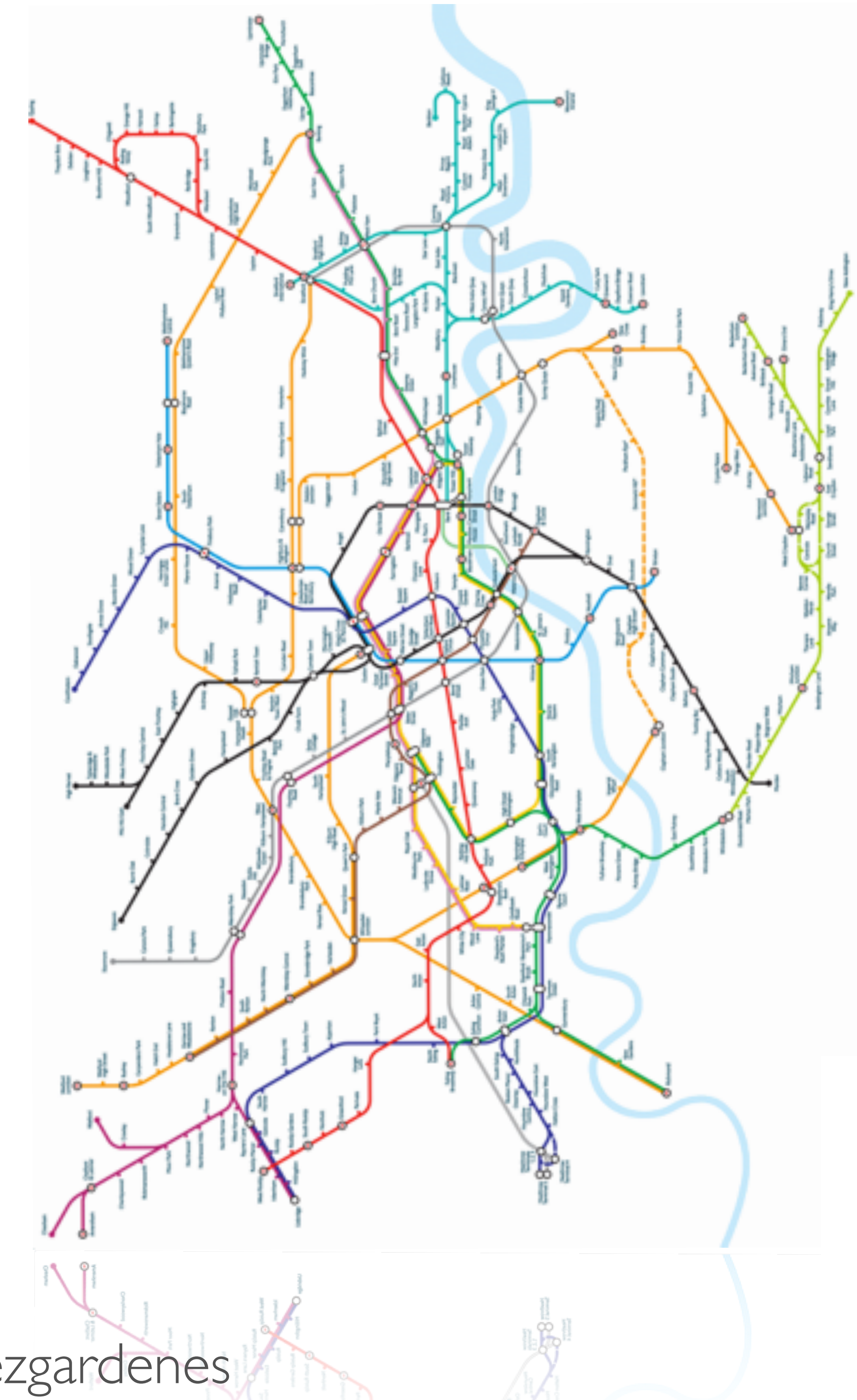


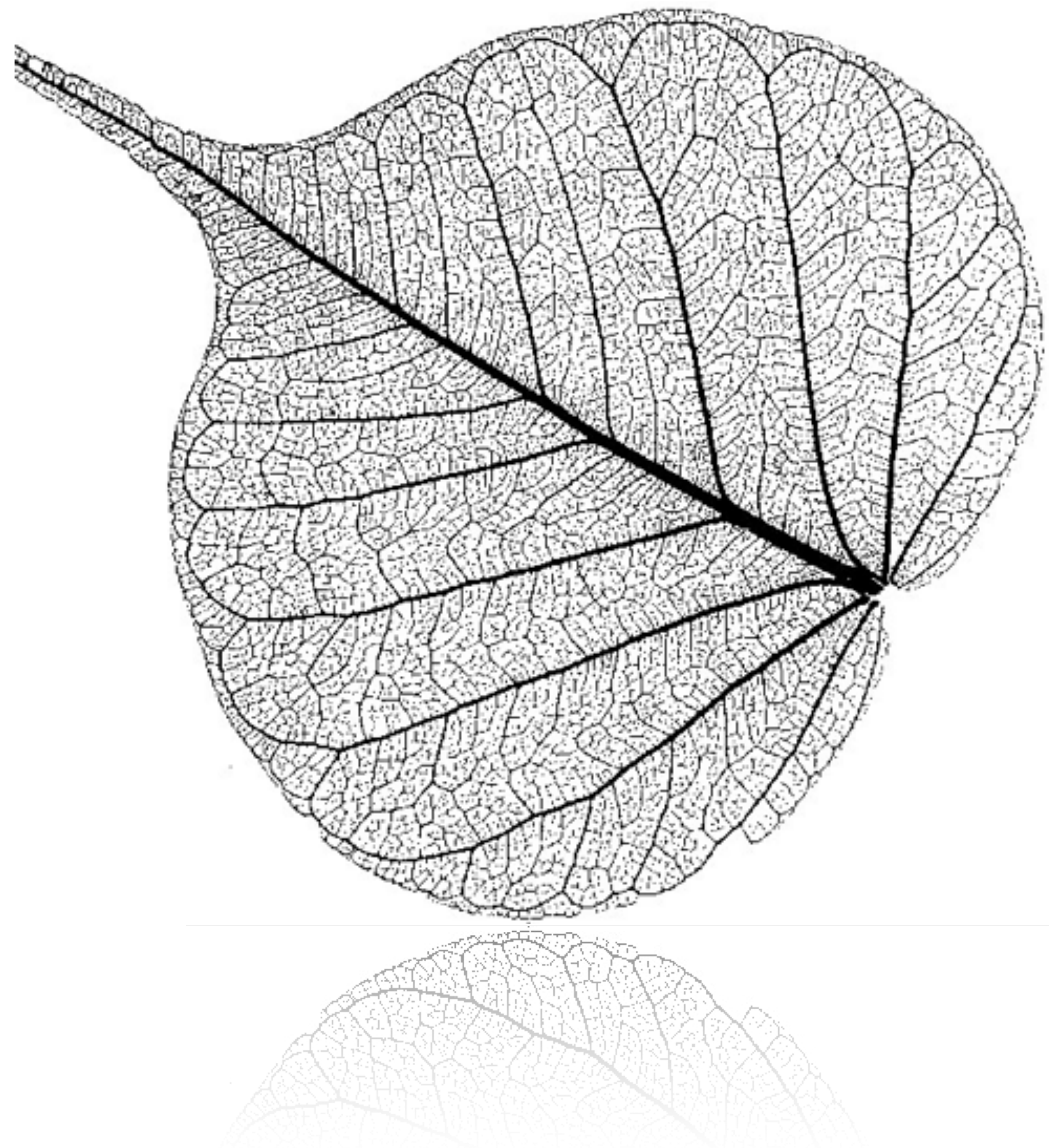




@gomezgardenes

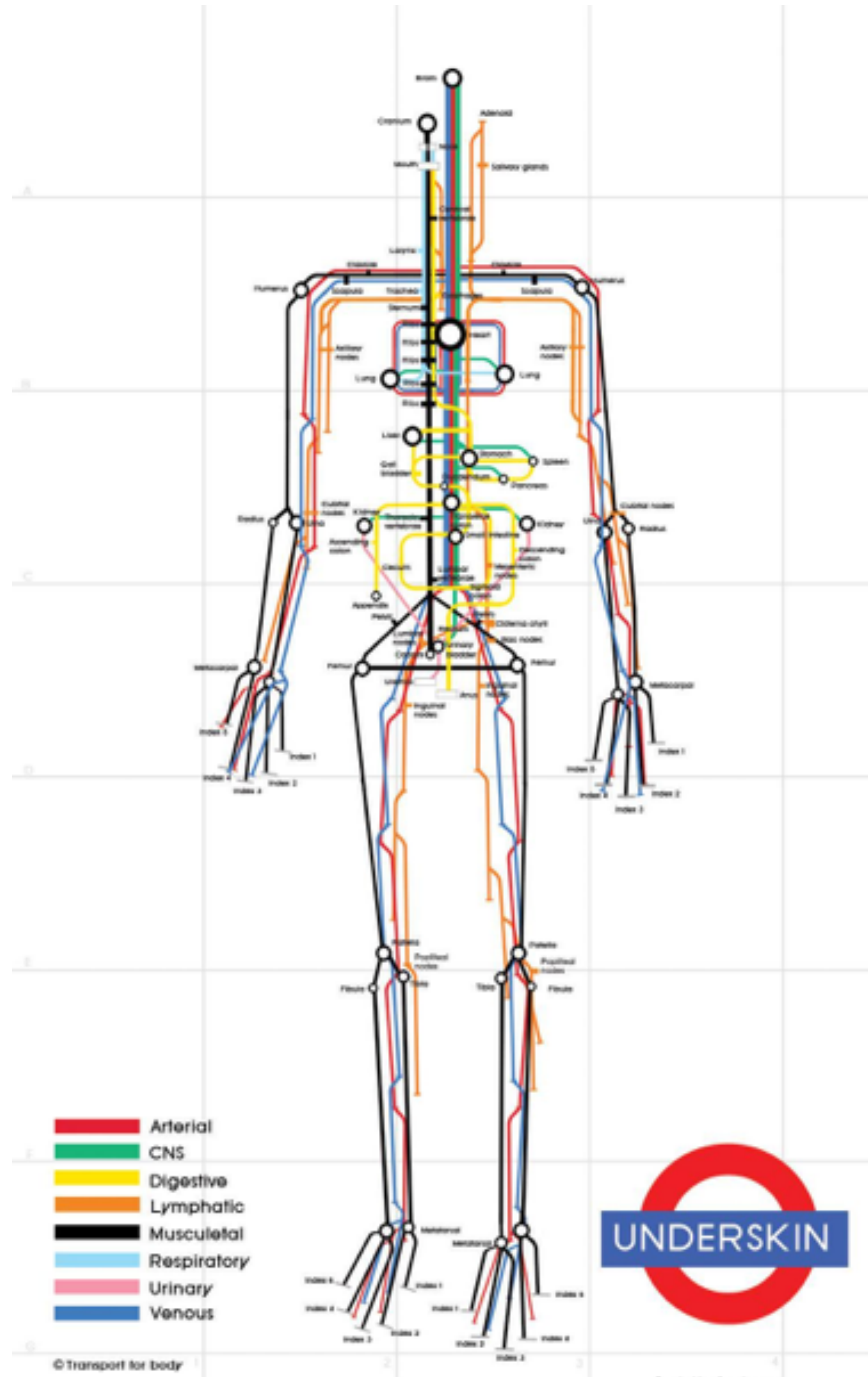
Goal: To reveal Similar (Universal) organizational principles of Complex Systems



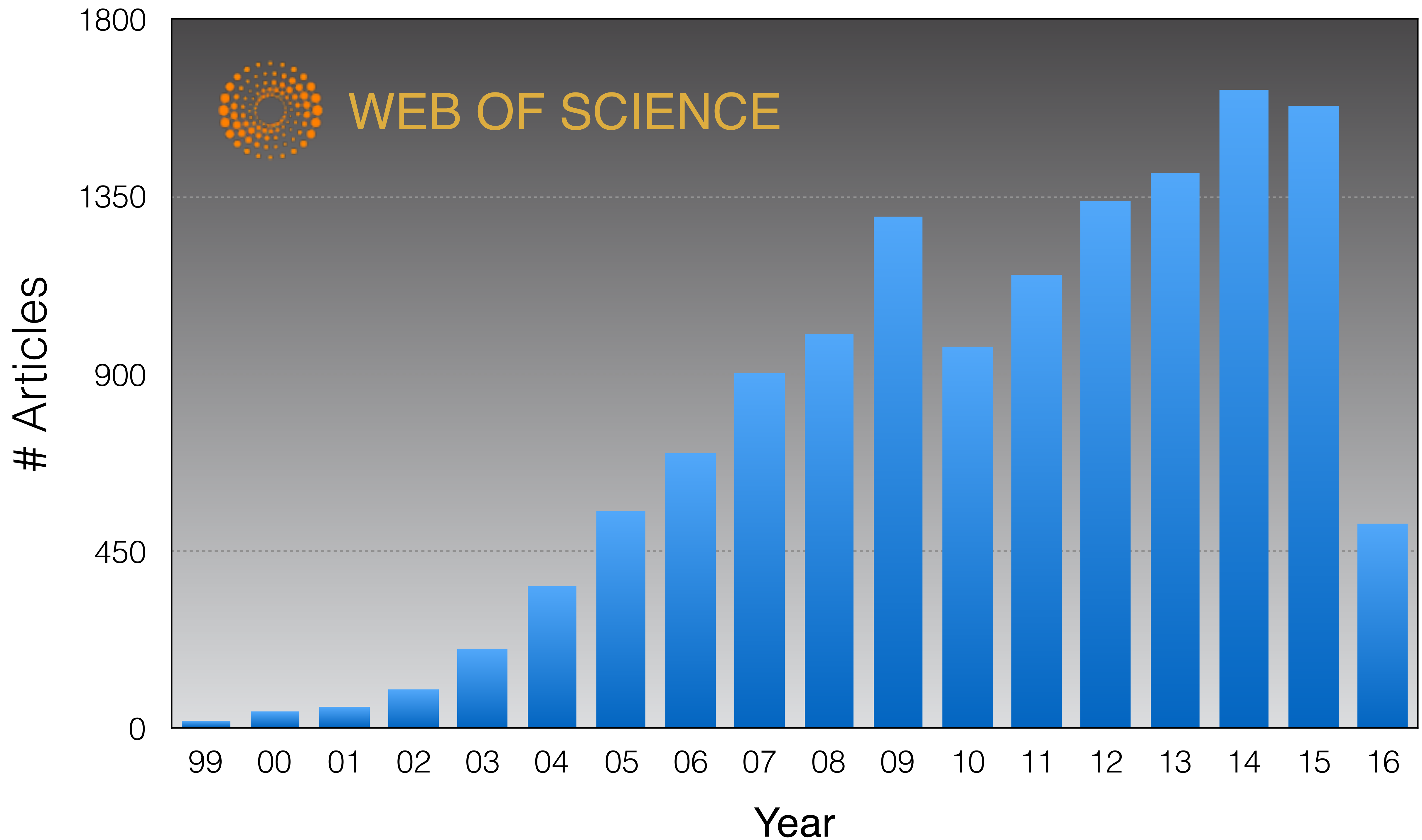


@gomezgardenes

Goal: To reveal Similar (Universal) organizational principles of Complex Systems



A growing field…



@gomezgardenes

Some Master References

Reviews

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Published 30 January 2002)

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256

The Structure and Function of Complex Networks*

M. E. J. Newman[†]

Advances in Physics **51** 1079 (2002)

Evolution of networks

S.N. Dorogovtsev[1,2,*] and J.F.F. Mendes[1,†]
[1] *Departamento de Física and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto Rua do Campo Alegre 687, 4169-007 Porto, Portugal*
[2] *A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*

ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT
Physics Reports 424 (2006) 175–308
PHYSICS REPORTS
www.elsevier.com/locate/physrep

Complex networks: Structure and dynamics

S. Boccaletti[a,*], V. Latora[b,c], Y. Moreno[d,e], M. Chavez[f], D.-U. Hwang[a]

Some Master References

Reviews

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Published 30 January 2002)

16774 citations

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256

© 2003 Society for Industrial and Applied Mathematics

The Structure and Function of Complex Networks*

M. E. J. Newman†

14024 citations

Advances in Physics **51** 1079 (2002)

Evolution of networks

S.N. Dorogovtsev[1,2,*] and J.F.F. Mendes[1,†]

[1] *Departamento de Física and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

[2] *A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*

3014 citations

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

PHYSICS REPORTS

Physics Reports 424 (2006) 175–308

www.elsevier.com/locate/physrep

Complex networks: Structure and dynamics

S. Boccaletti[a,*], V. Latora[b,c], Y. Moreno[d,e], M. Chavez[f], D.-U. Hwang[a]

6324 citations

@gomezgardenes

Some Master References

Reviews

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

(Published 30 January 2002)

16774 citations

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256

The Structure and Function of Complex Networks*

M. E. J. Newman†

14024 citations

Advances in Physics **51** 1079 (2002)

Evolution of networks

S.N. Dorogovtsev[1,2,*] and J.F.F. Mendes[1,†]

[1] *Departamento de Física and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

[2] *A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*

3014 citations

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT

PHYSICS REPORTS

Physics Reports 424 (2006) 175–308

www.elsevier.com/locate/physrep

Complex networks: Structure and dynamics

S. Boccaletti[a,*], V. Latora[b,c], Y. Moreno[d,e], M. Chavez[f], D.-U. Hwang[a]

6324 citations

(some more to come)

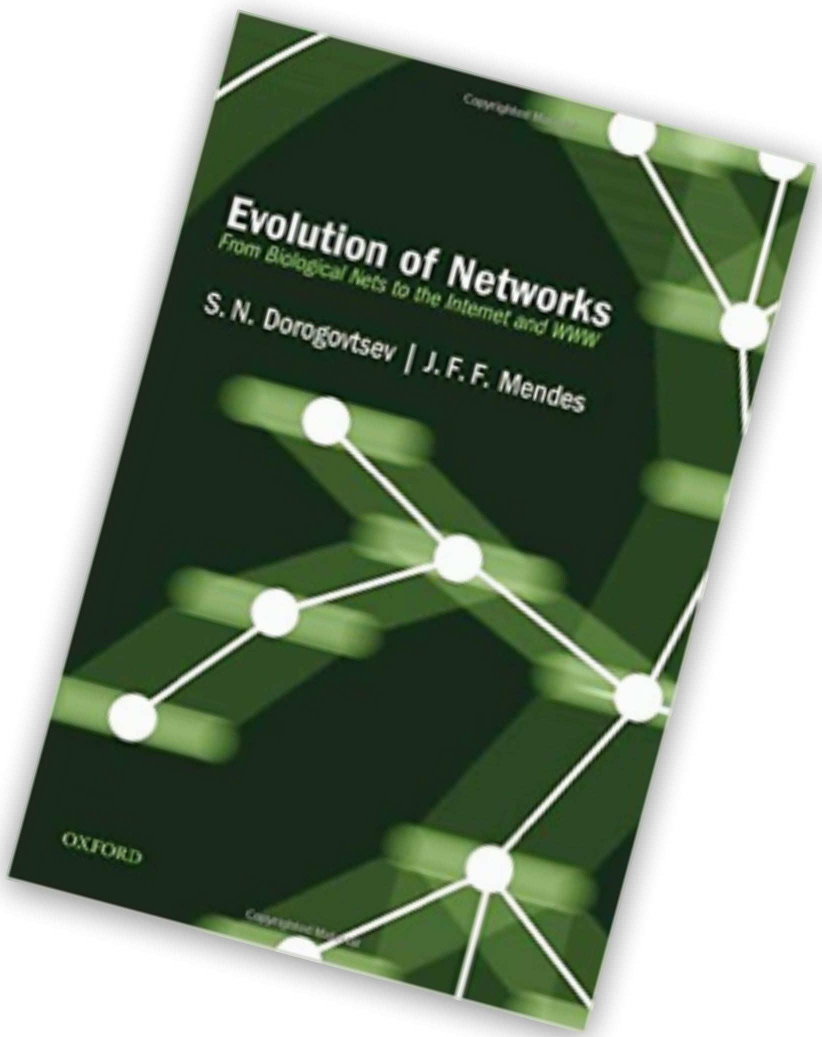
@gomezgardenes

Some Master References

Books

Some Master References

Books



2003

@gomezgardenes

Some Master References

Books

2003 2008

@gomezgardenes

Some Master References

Books

2003 2008 2010

@gomezgardenes

Some Master References

Books

2003 2008 2010 2011

@gomezgardenes

Some Master References

Books

2003

2008

2010

2011

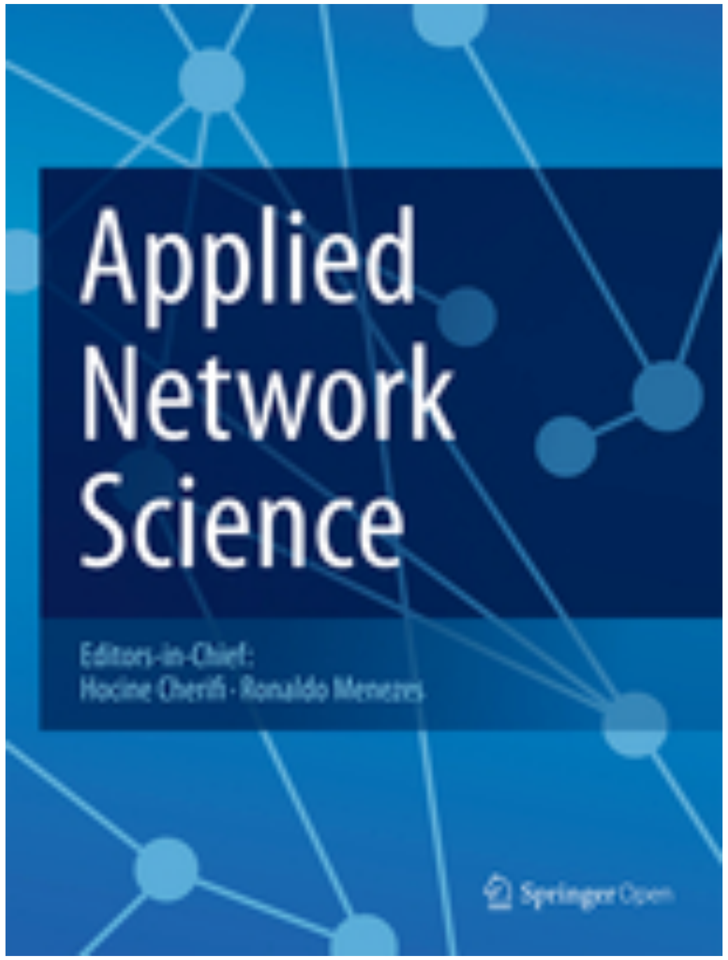
2016

@gomezgardenes

Specific Journals

2013

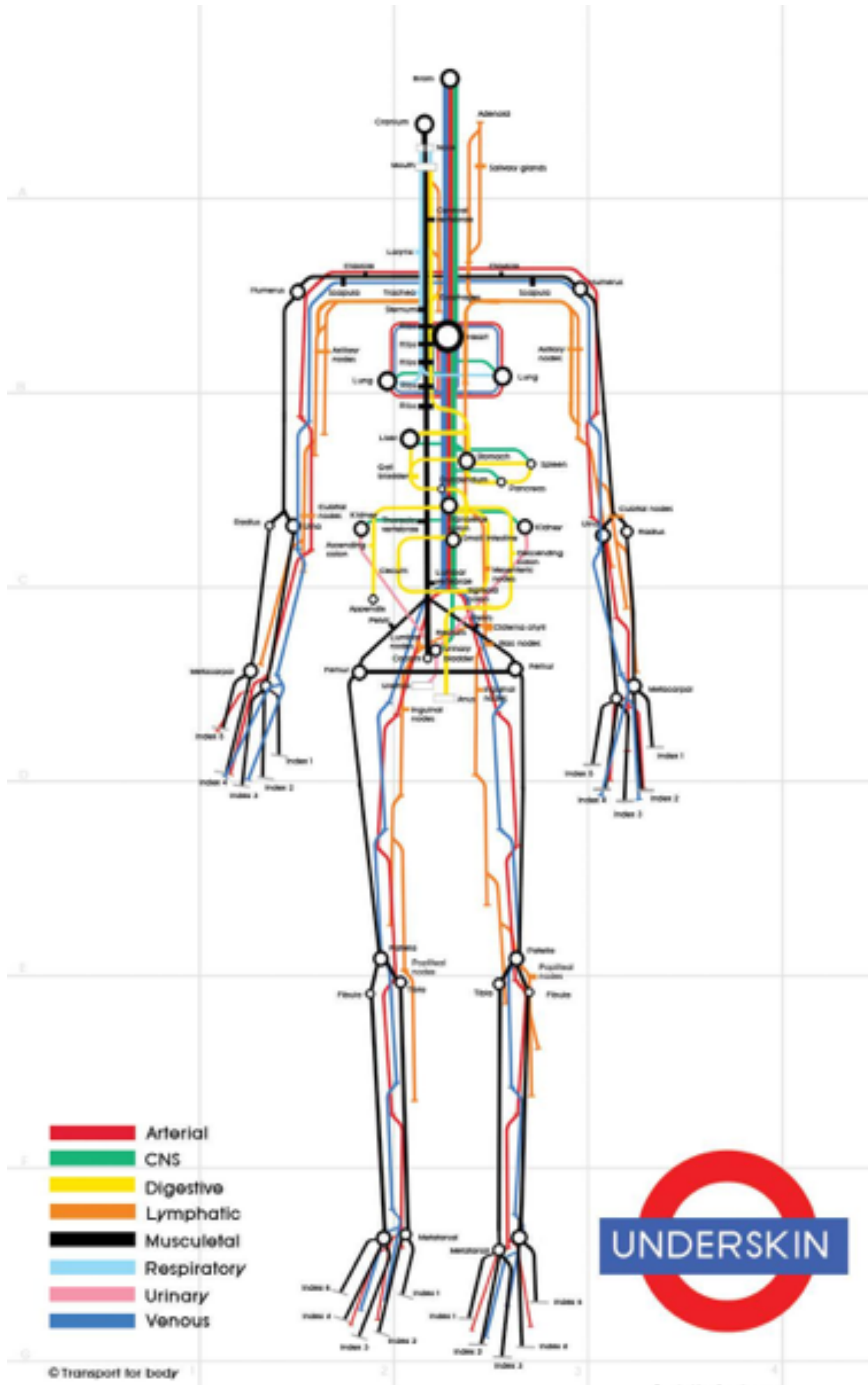
2013



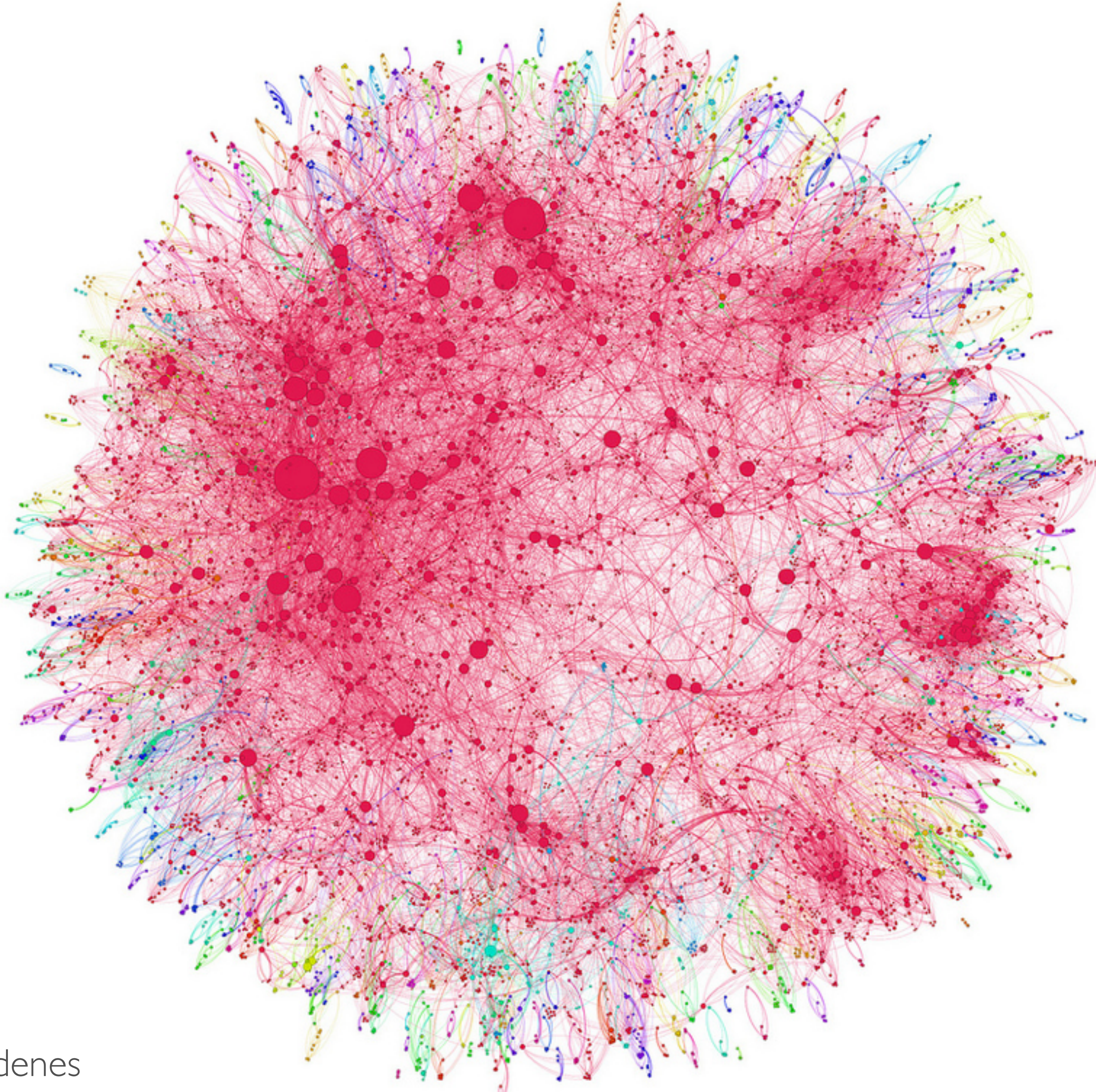
2016

@gomezgardenes

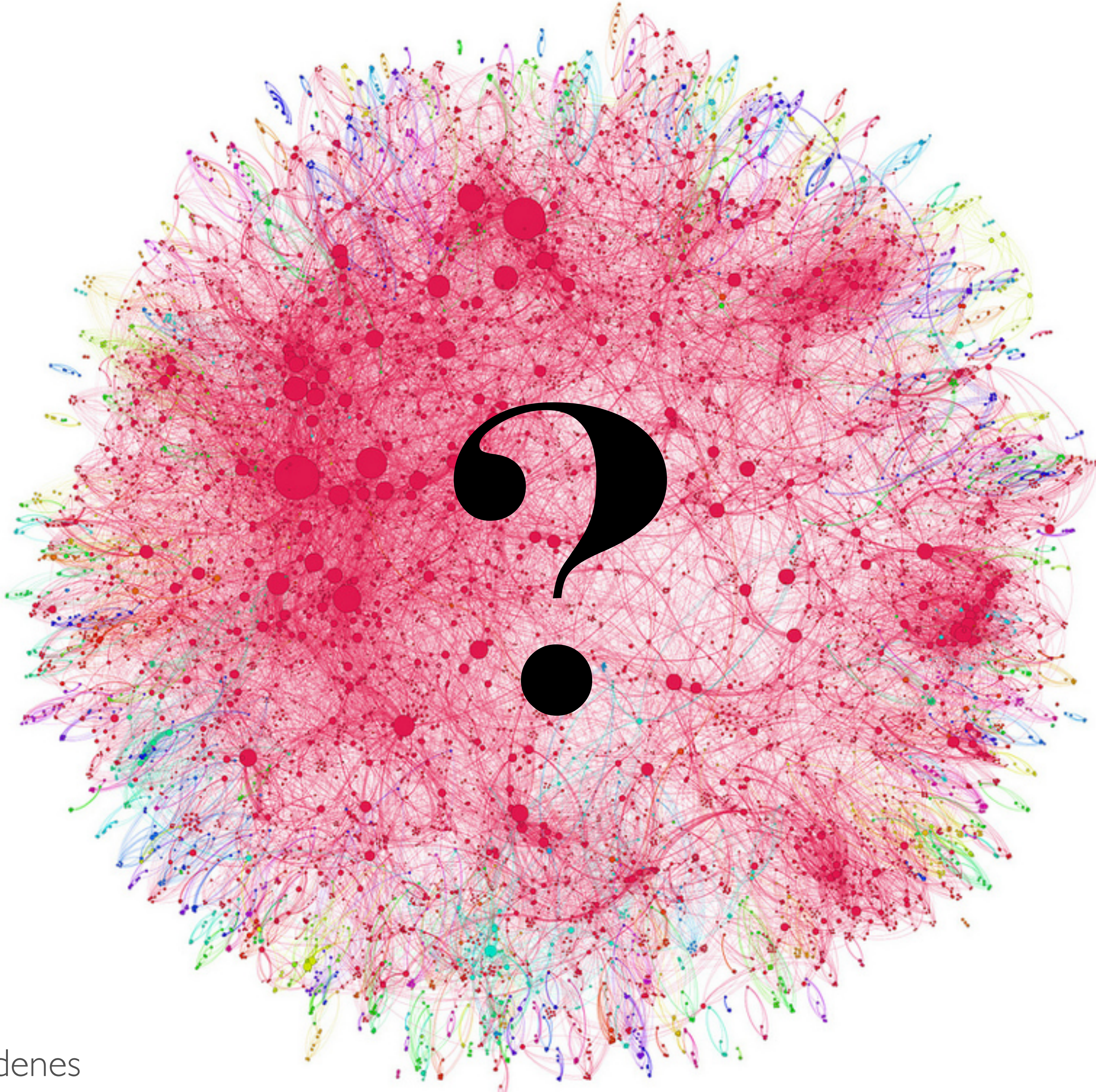
Goal: To reveal Similar (Universal) organizational principles of Complex Systems



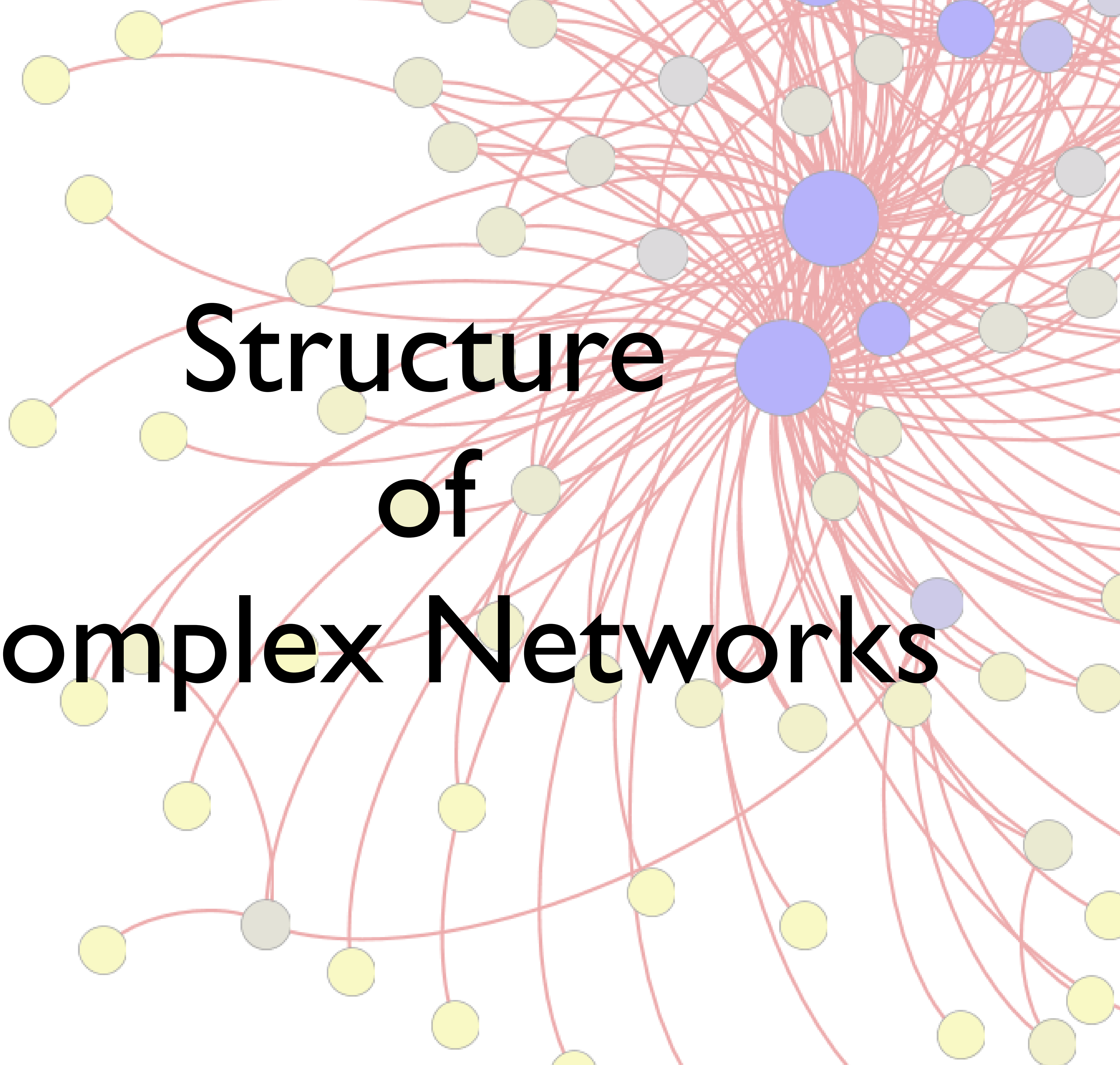
@gomezgardenes



@gomezgardenes



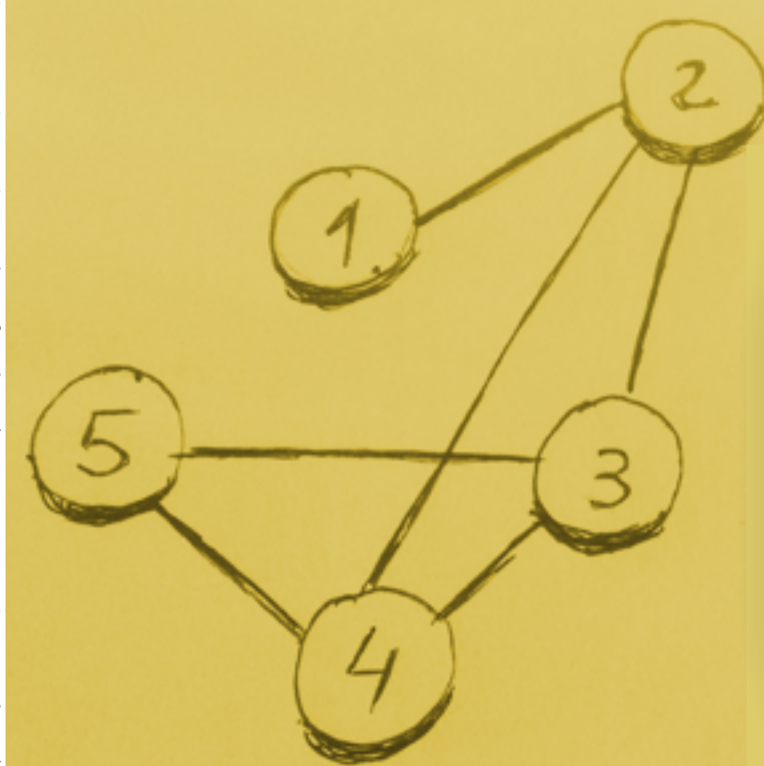
@gomezgardenes



Structure of Complex Networks

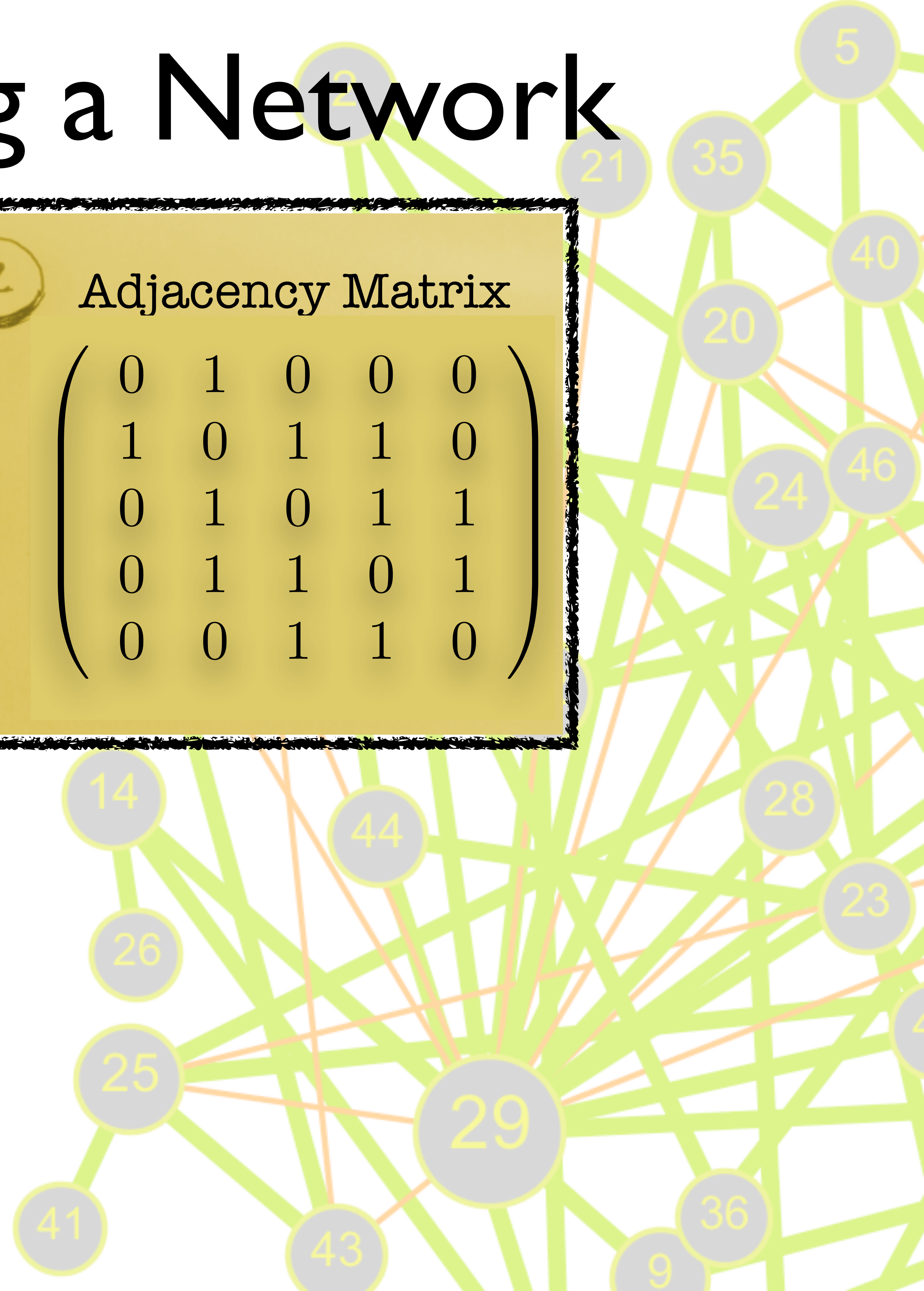
@gomezgardenes

Encoding a Network

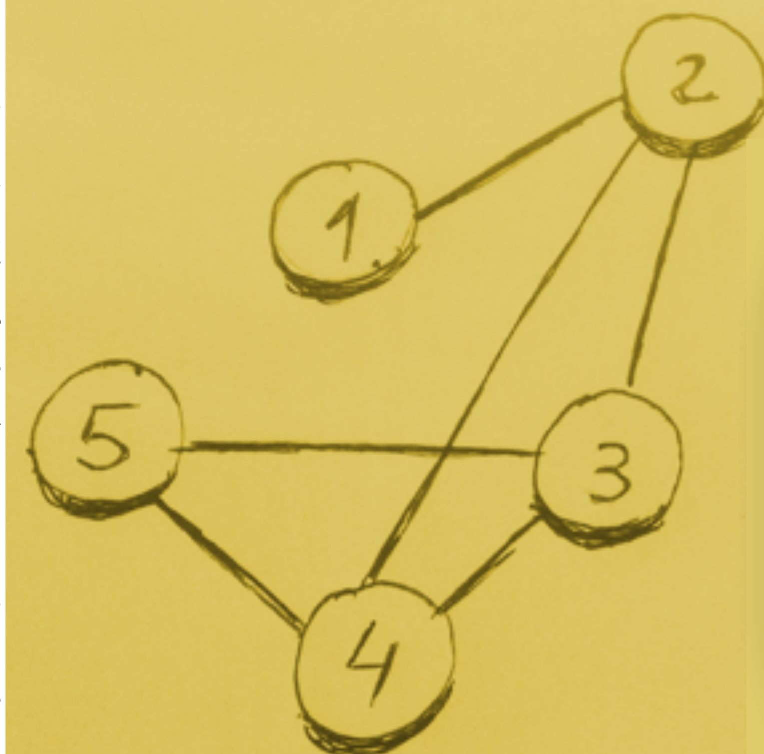


Adjacency Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$



Encoding a Network



Adjacency Matrix

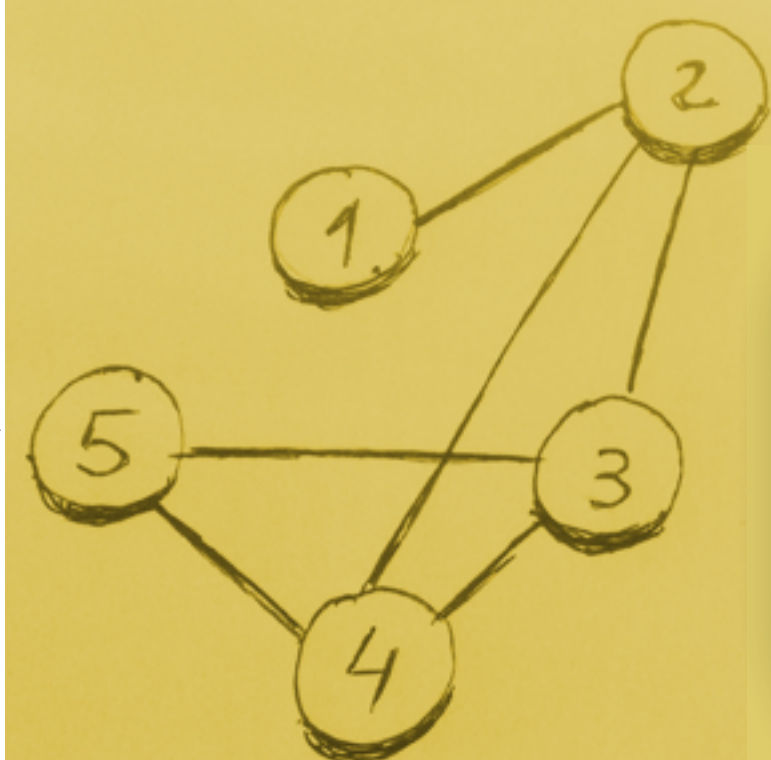
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Network Laplacian

$$\mathcal{L} = \begin{pmatrix} k_1 & 0 & . & 0 & 0 \\ 0 & k_2 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & k_{N-1} & 0 \\ 0 & 0 & . & 0 & k_N \end{pmatrix} - A$$

@gomezgardenes

Encoding a Network



Adjacency Matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Network Laplacian

$$\mathcal{L} = \begin{pmatrix} k_1 & 0 & . & 0 & 0 \\ 0 & k_2 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & k_{N-1} & 0 \\ 0 & 0 & . & 0 & k_N \end{pmatrix} - A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 3 & -1 & -1 \\ 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{pmatrix}$$

@gomezgardenes

Main Global Descriptors

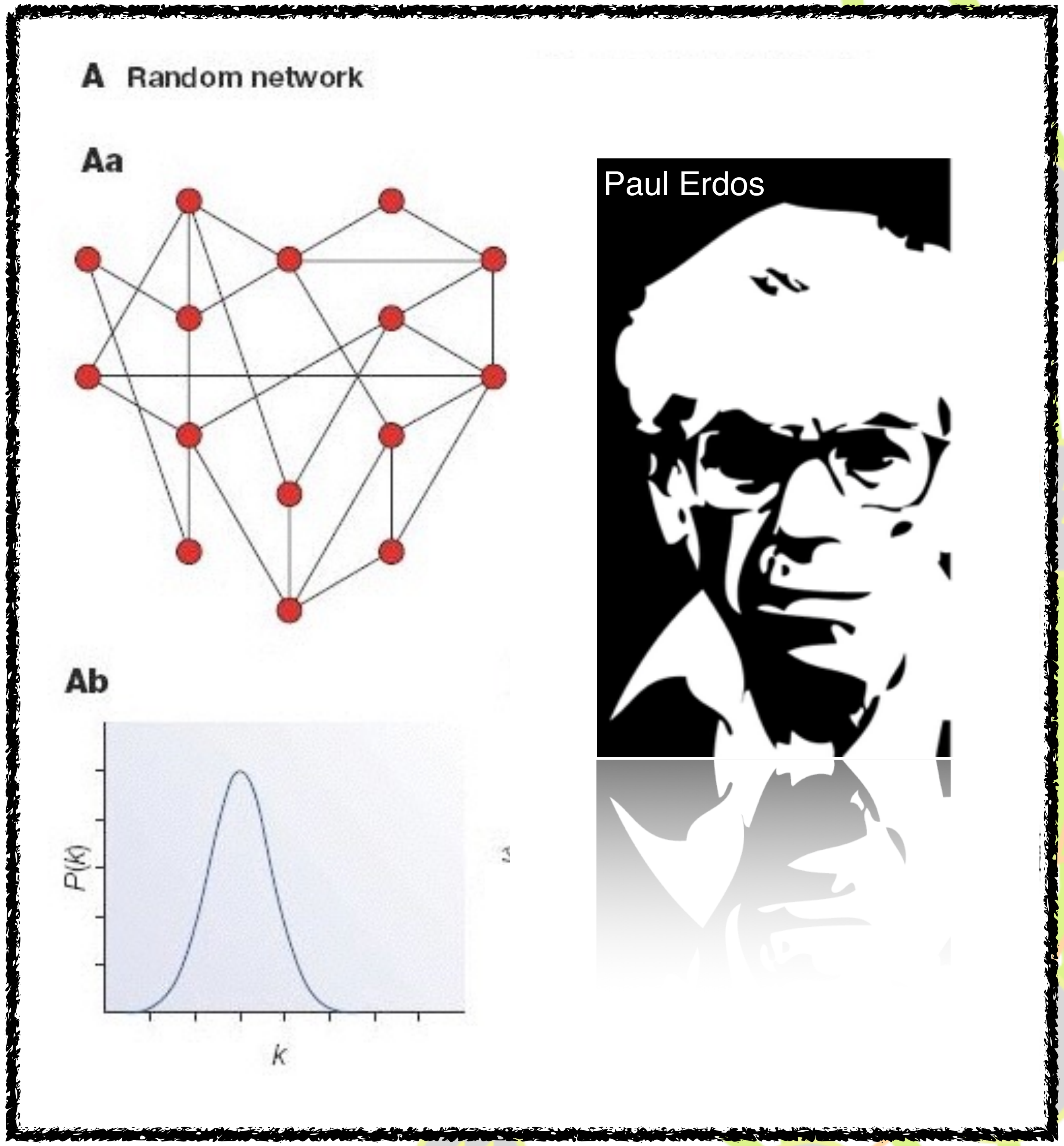
- Degree Distribution
- Clustering Coefficient
- Distances
- Correlations
- Centrality
- K-Cores
- Motifs
- Communities
- ....

Main Global Descriptors

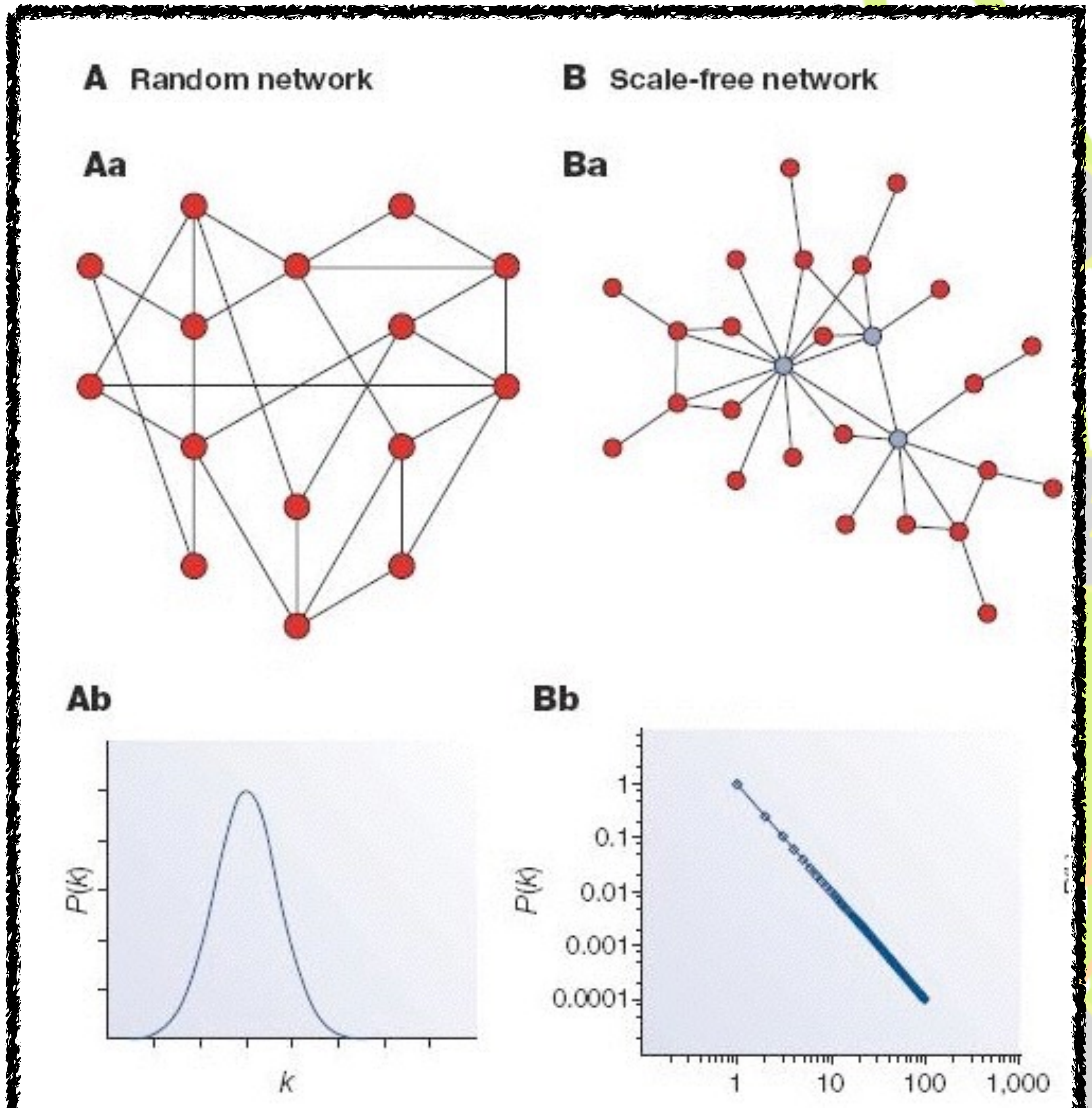
- Degree Distribution
- Clustering Coefficient
- Distances
- Correlations
- Centrality
- K-Cores
- Motifs
- Communities
- ....

@gomezgardenes

I Degree Distribution



I Degree Distribution

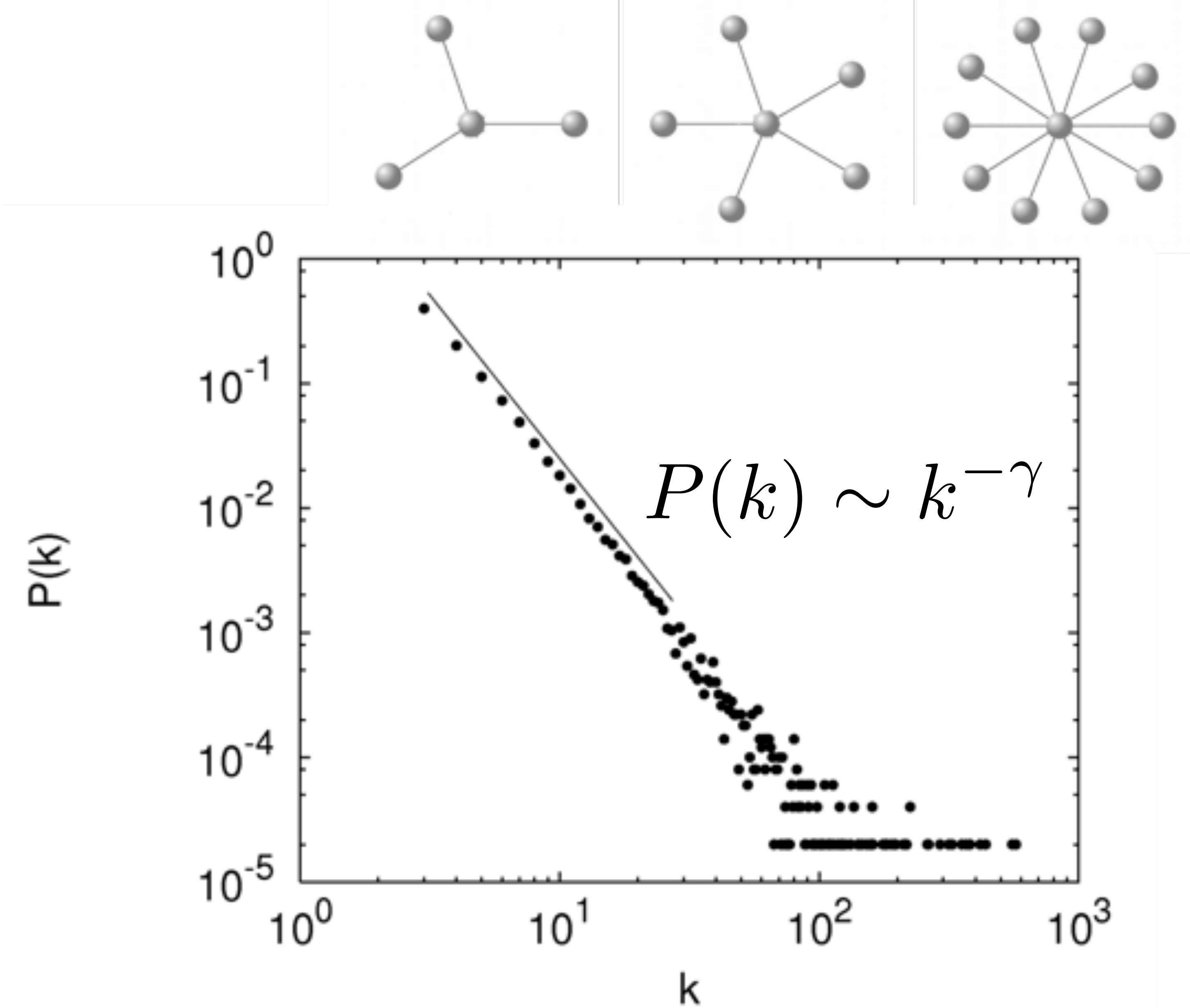


JEONG ET AL. NATURE 407, 651 (2000)

@gomezgardenes

I Degree Distribution

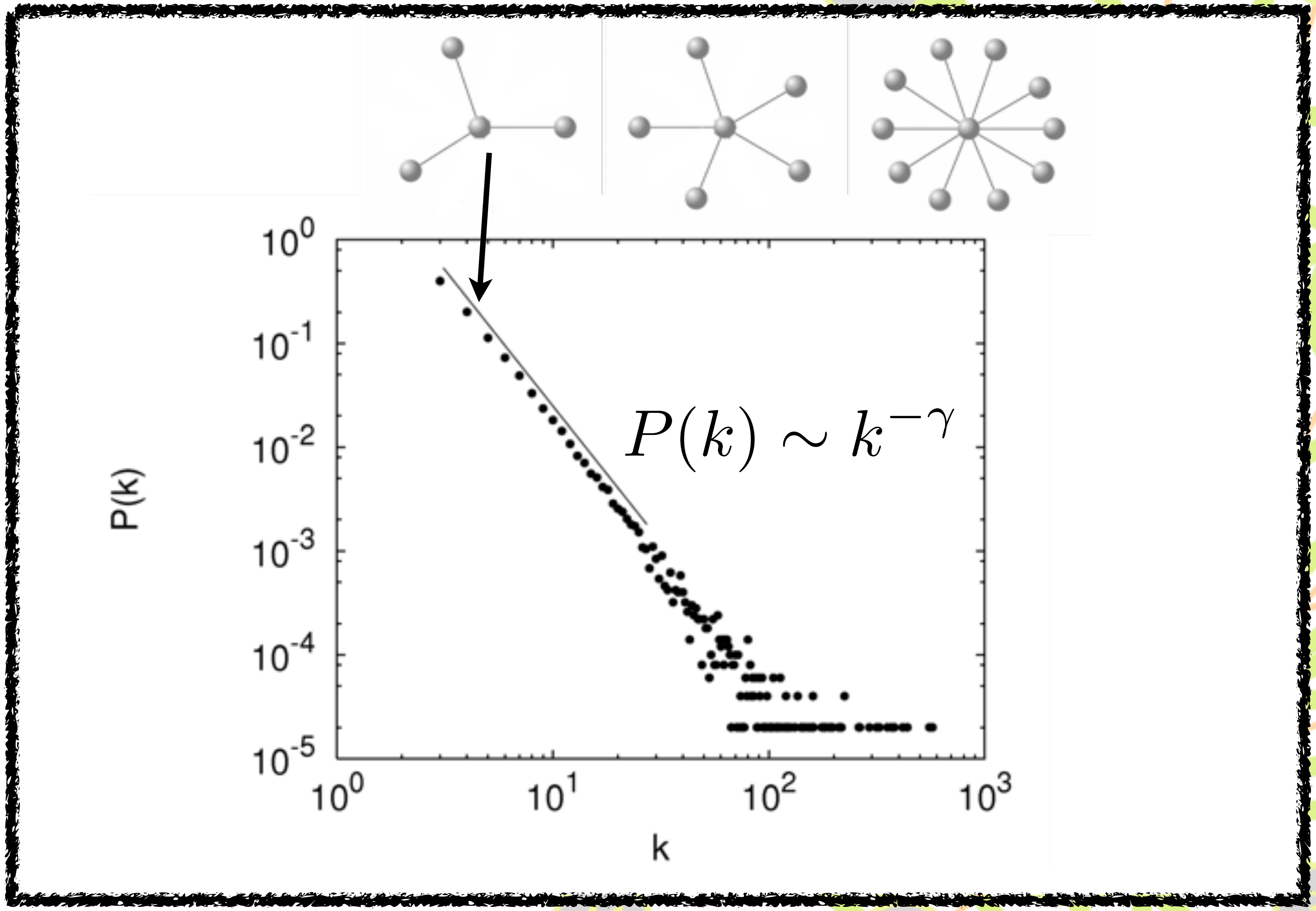
Scale-free phenomenon



@gomezgardenes

I Degree Distribution

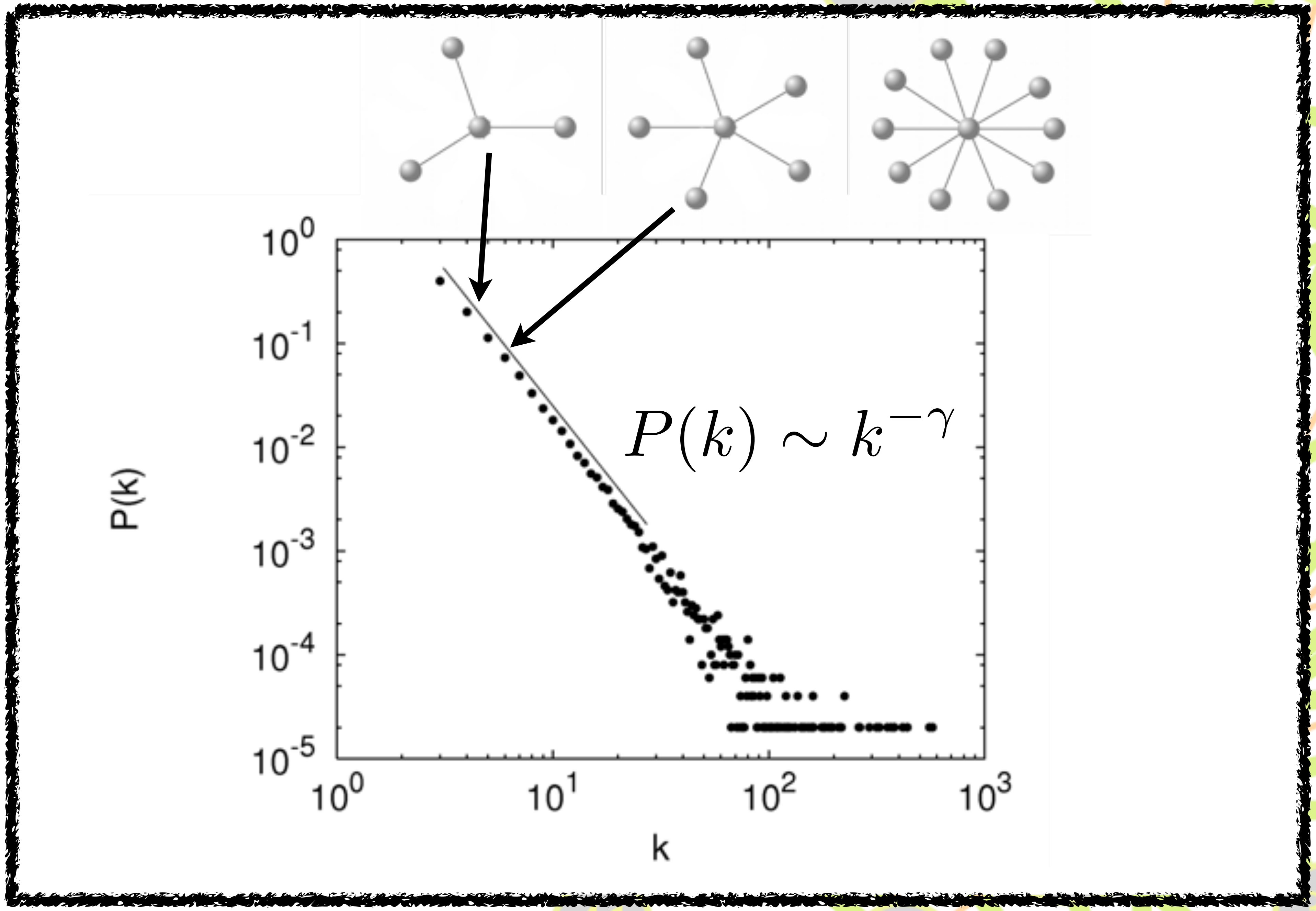
Scale-free phenomenon



@gomezgardenes

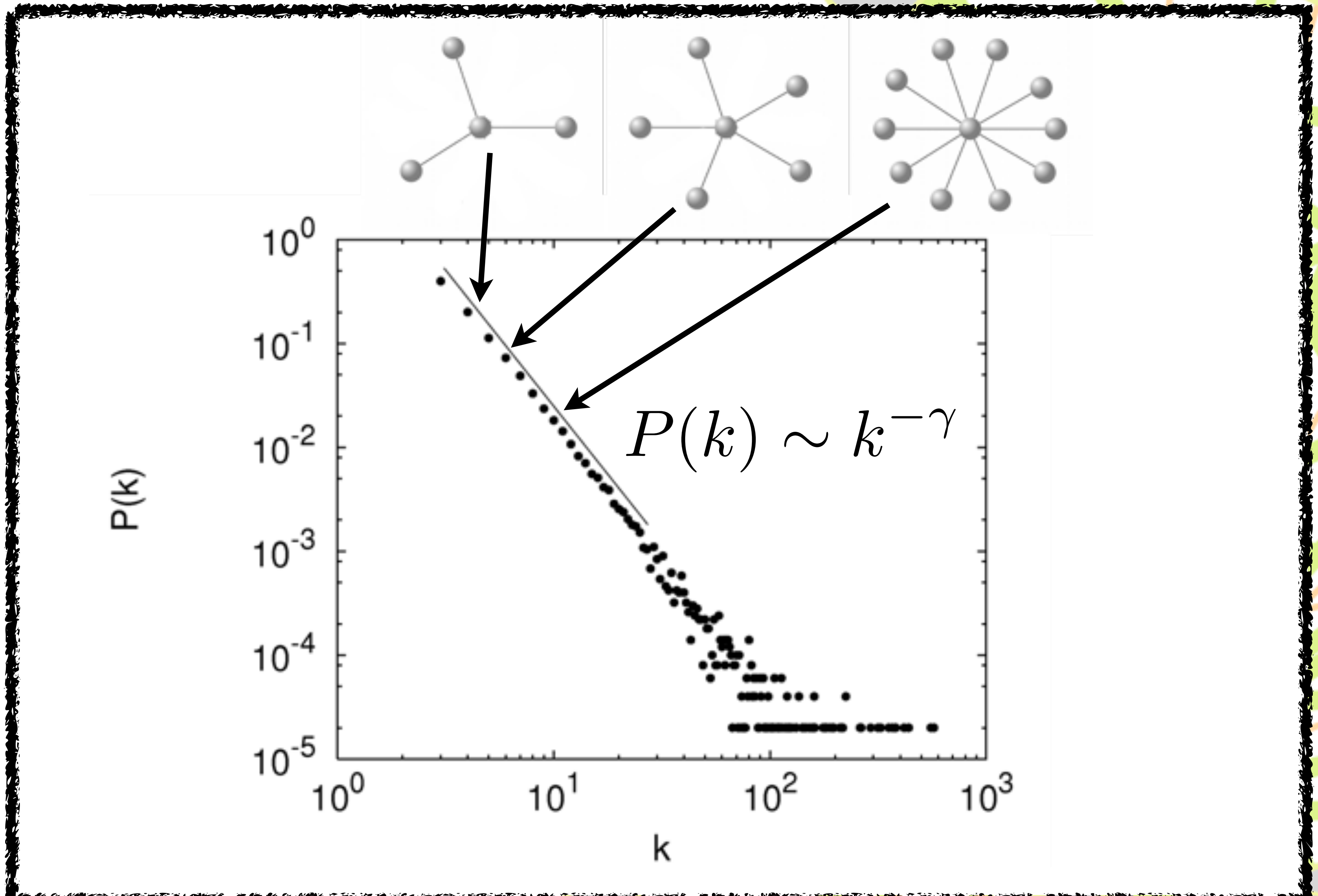
I Degree Distribution

Scale-free phenomenon



I Degree Distribution

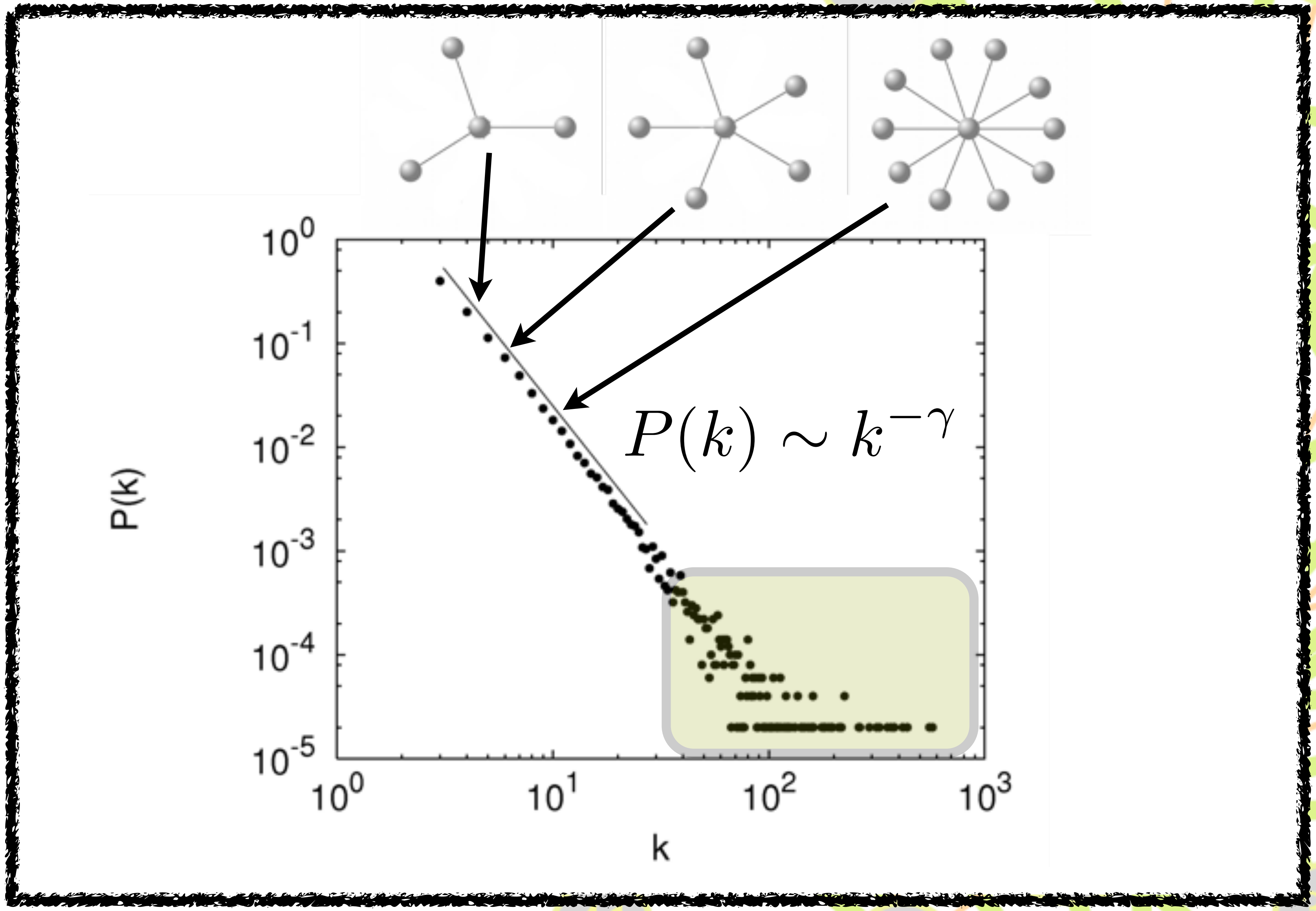
Scale-free phenomenon



@gomezgardenes

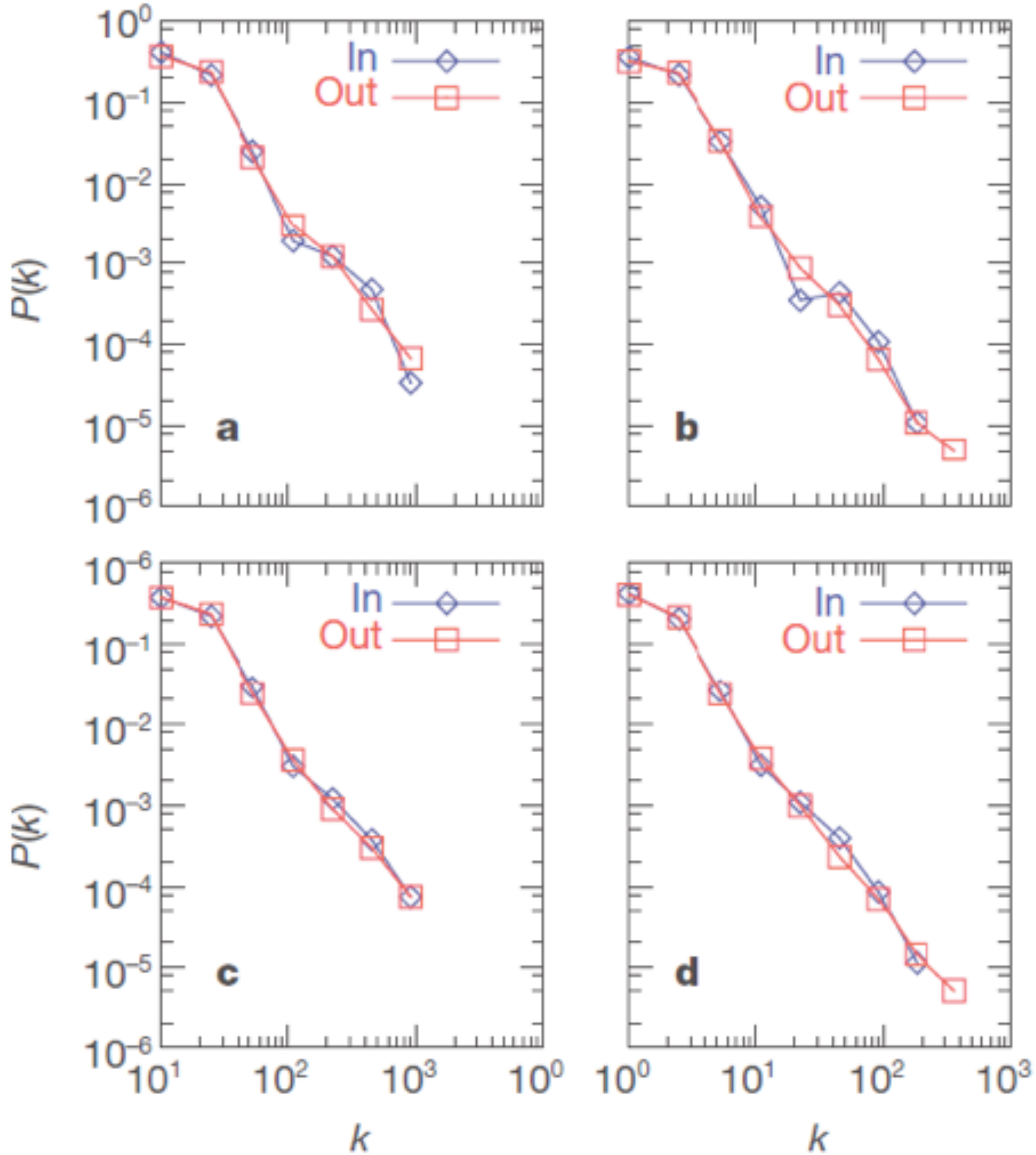
I Degree Distribution

Scale-free phenomenon



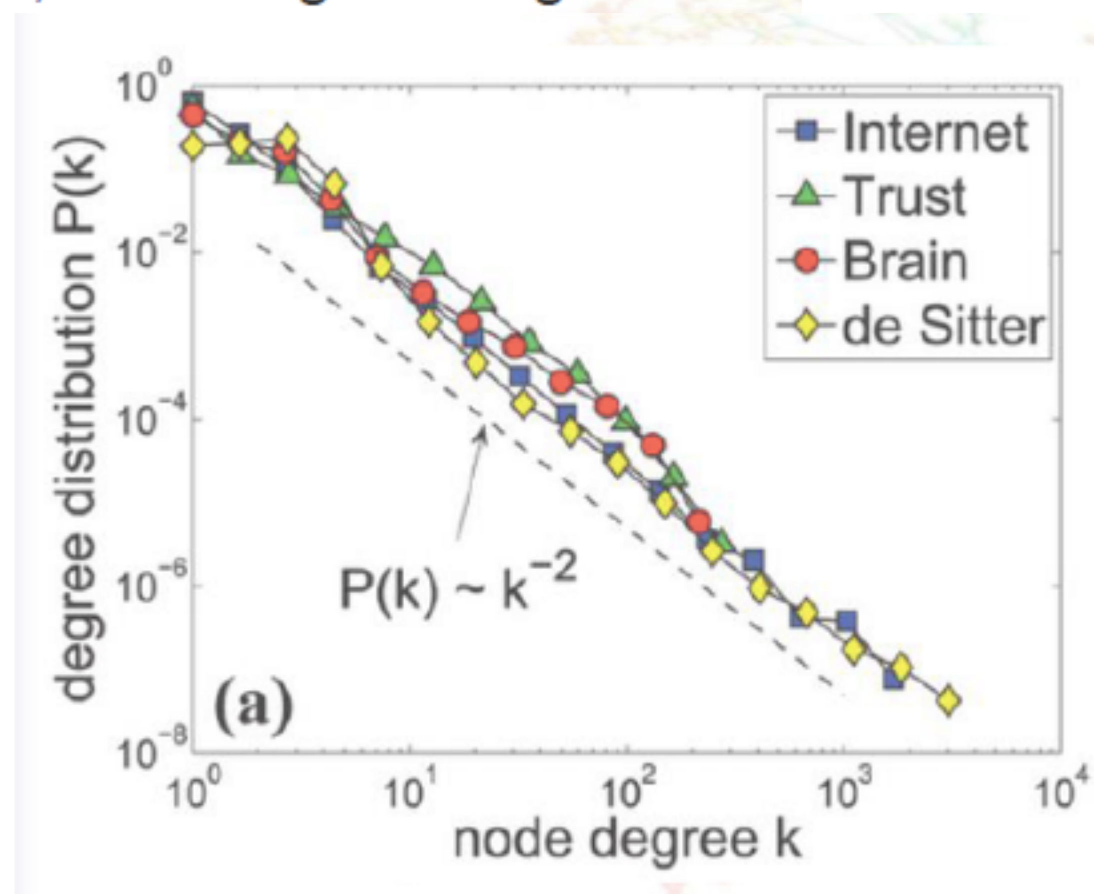
@gomezgardenes

I Degree Distribution



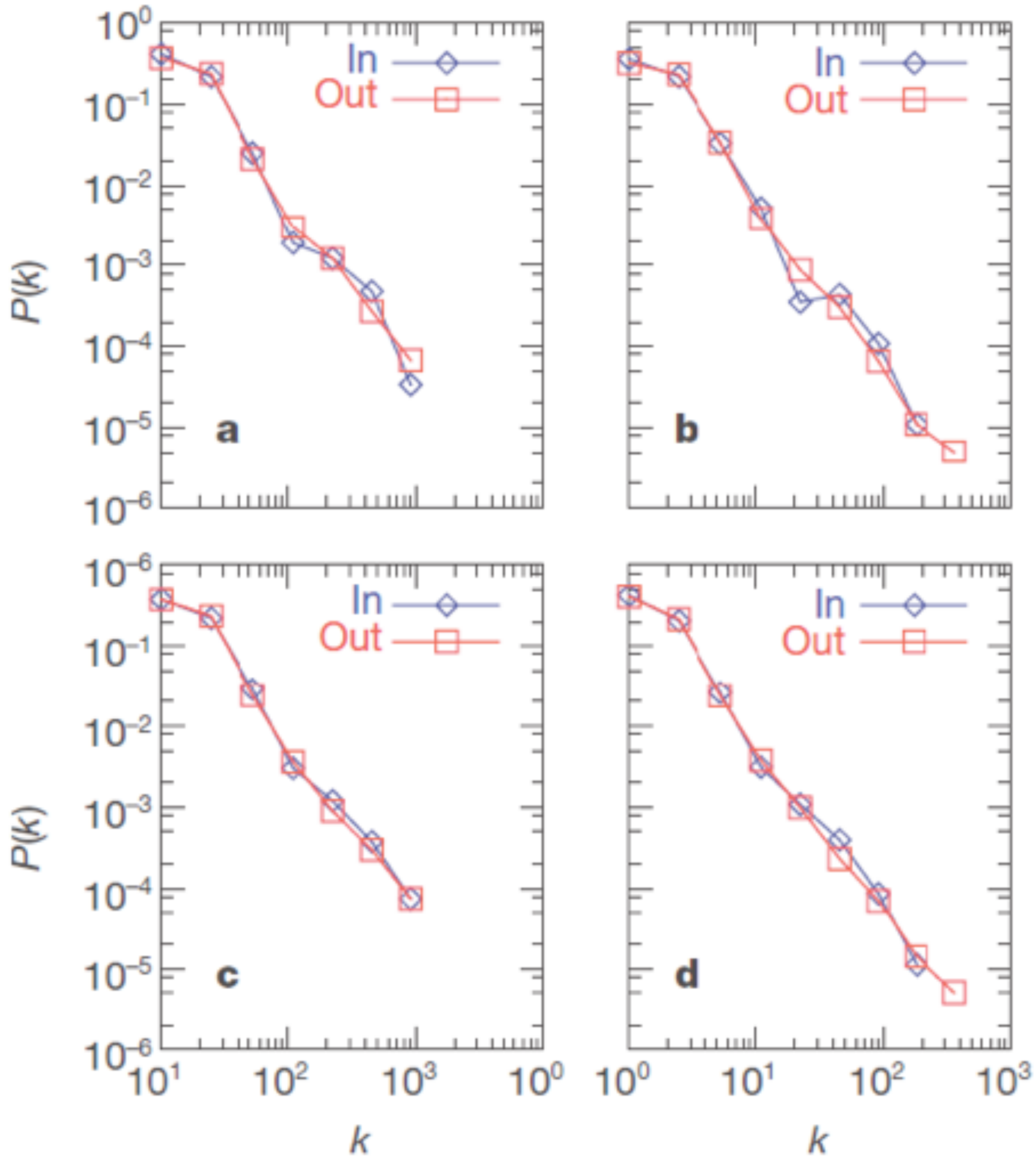
JEONG ET AL. NATURE 407, 651 (2000)

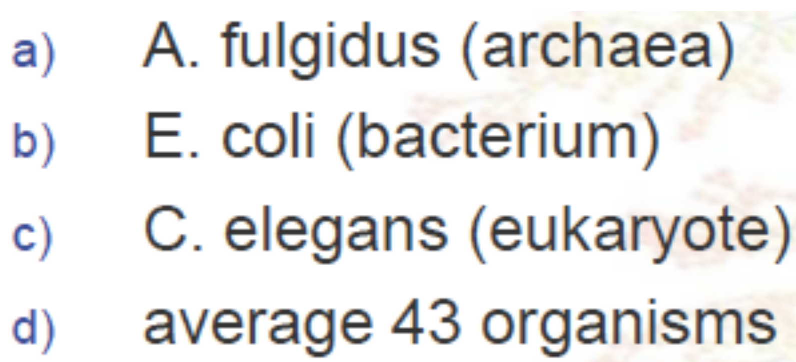
a) A. fulgidus (archaea)
b) E. coli (bacterium)
c) C. elegans (eukaryote)
d) average 43 organisms



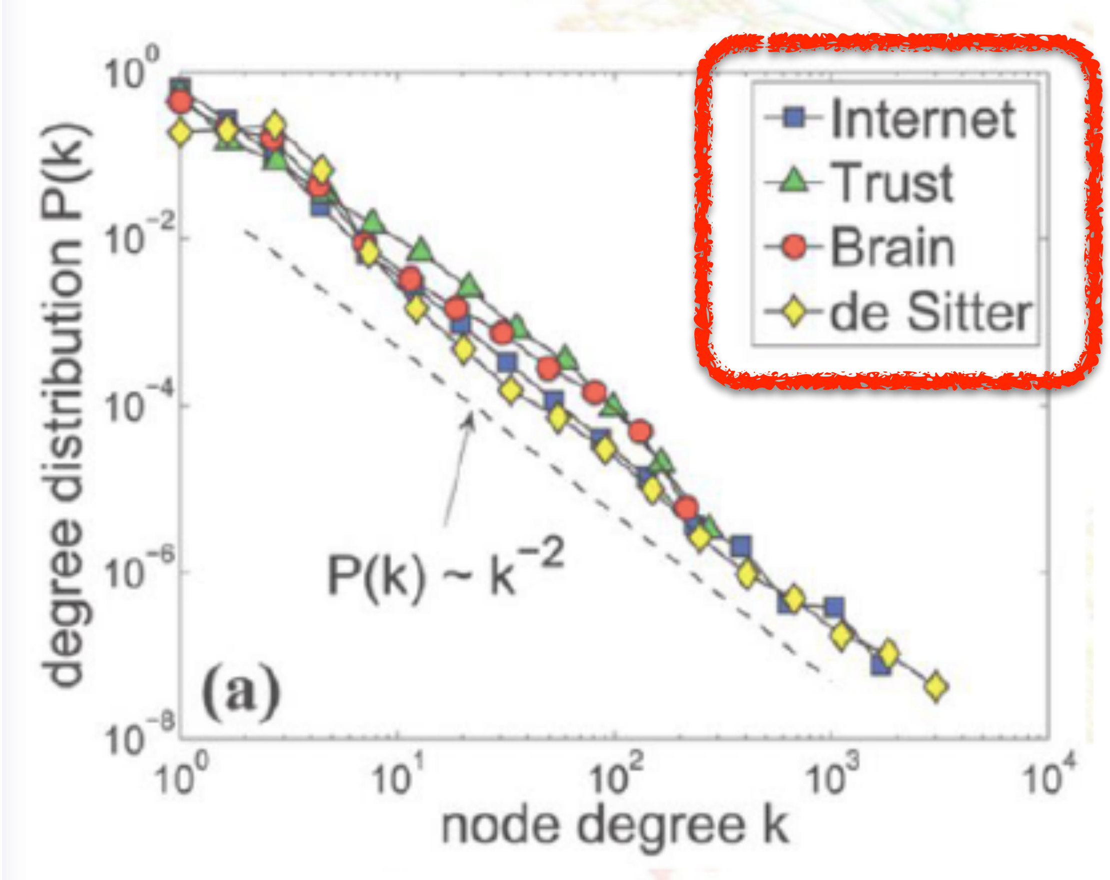
KRIOUKOV ET AL. SCIENTIFIC REPORTS 2, 793 (2012)

I Degree Distribution





JEONG ET AL. NATURE 407, 651 (2000)



KRIOUKOV ET AL. SCIENTIFIC REPORTS 2, 793 (2012)

@gomezgardenes

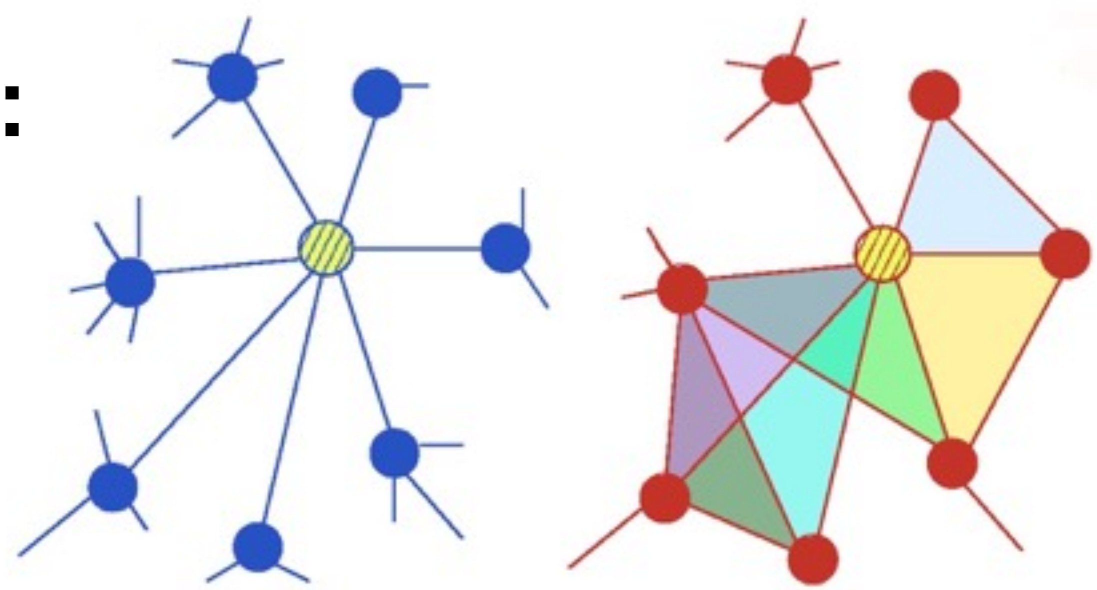
II Clustering Coefficient

Clustering of a node:

$$C_i = \frac{\#\text{triangles connected to } i}{\#\text{possible triangles connected to } i} = \frac{2E_i}{k_i(k_i-1)}$$

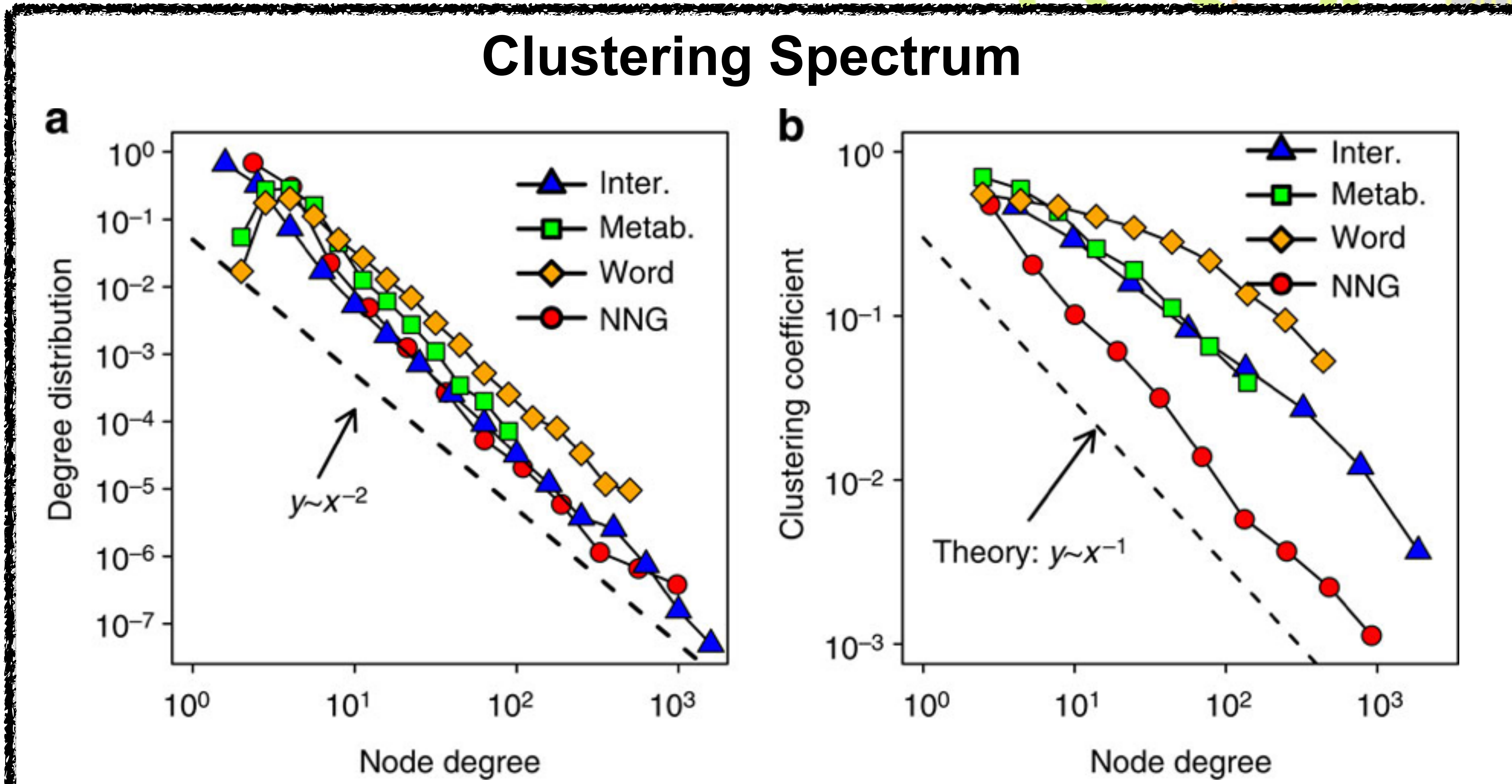
Clustering of the Network:

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i$$



@gomezgardenes

II Clustering Coefficient

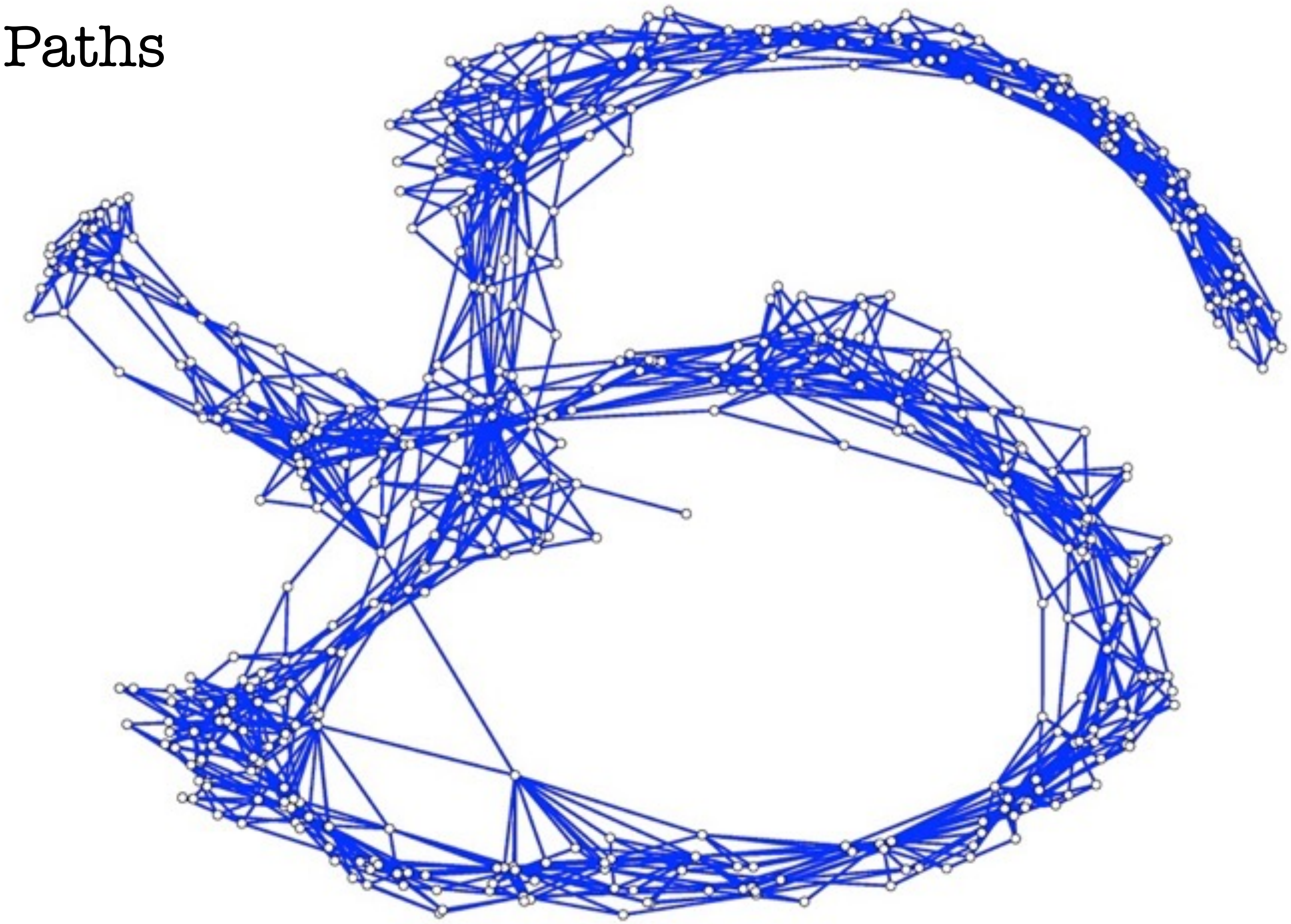


GULYÁS ET AL. NATURE COMMUNICATIONS 6, 7651 (2015)

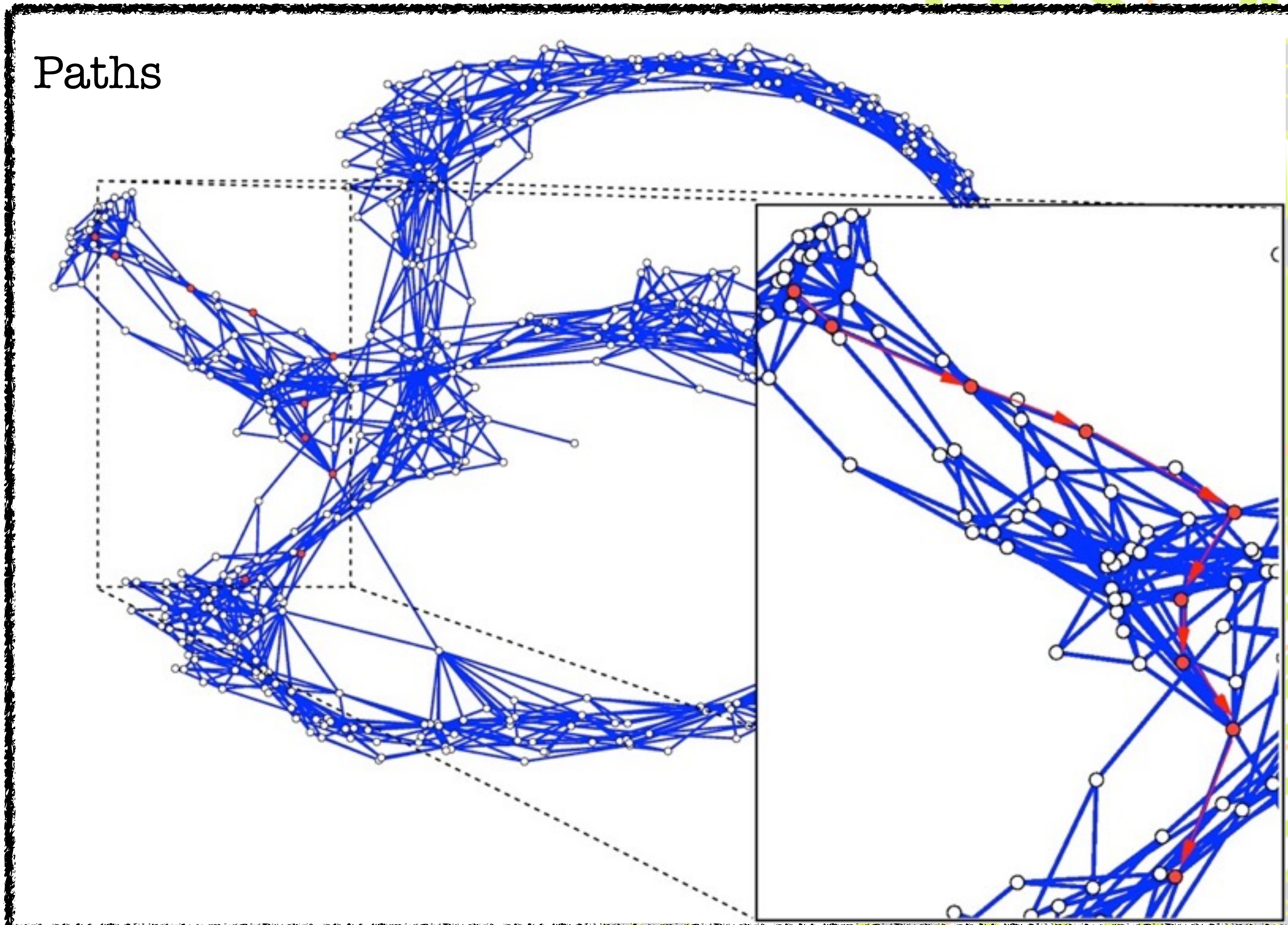
@gomezgardenes

III Distances

Paths



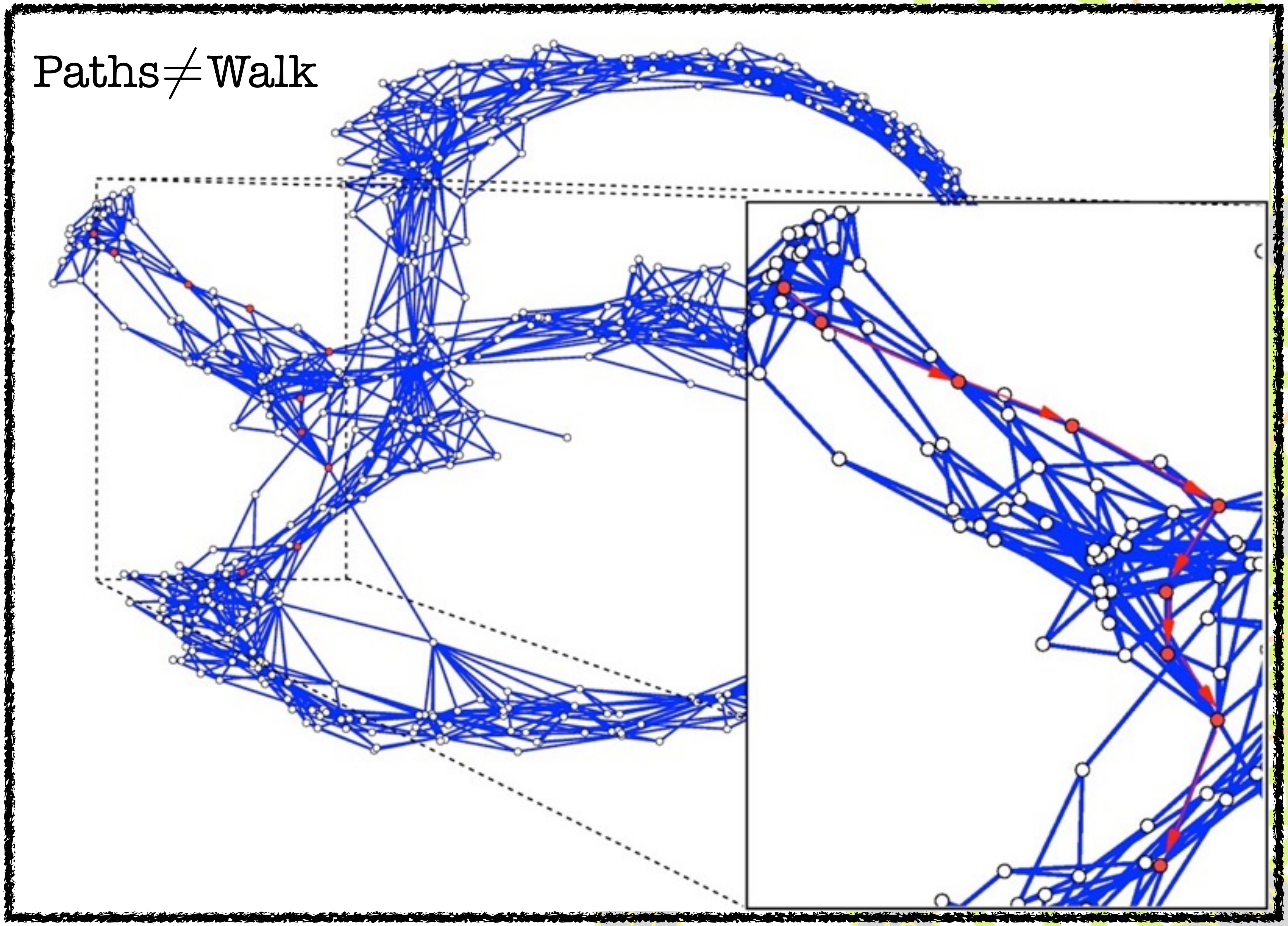
III Distances



@gomezgardenes

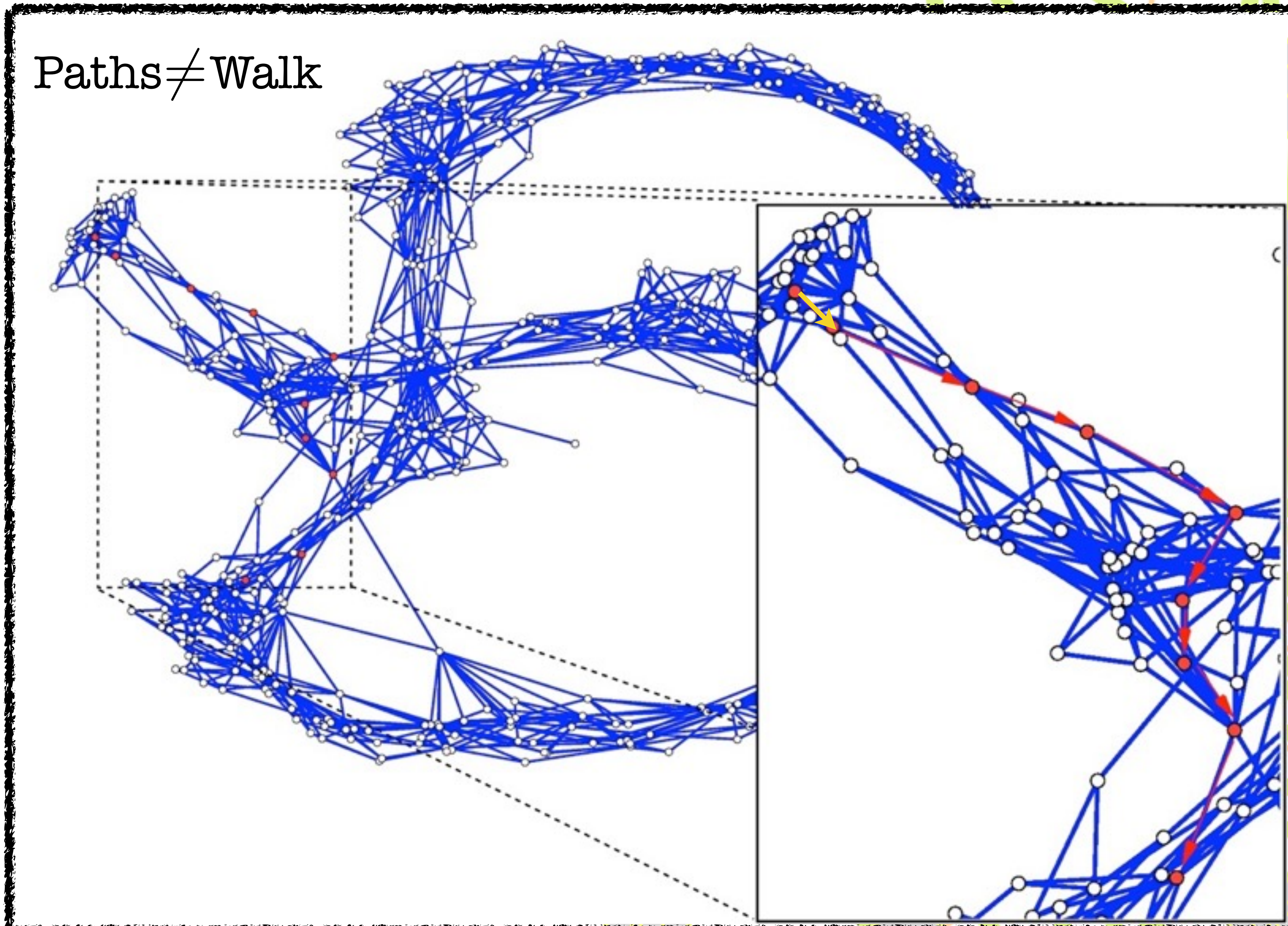
III Distances

Paths≠Walk

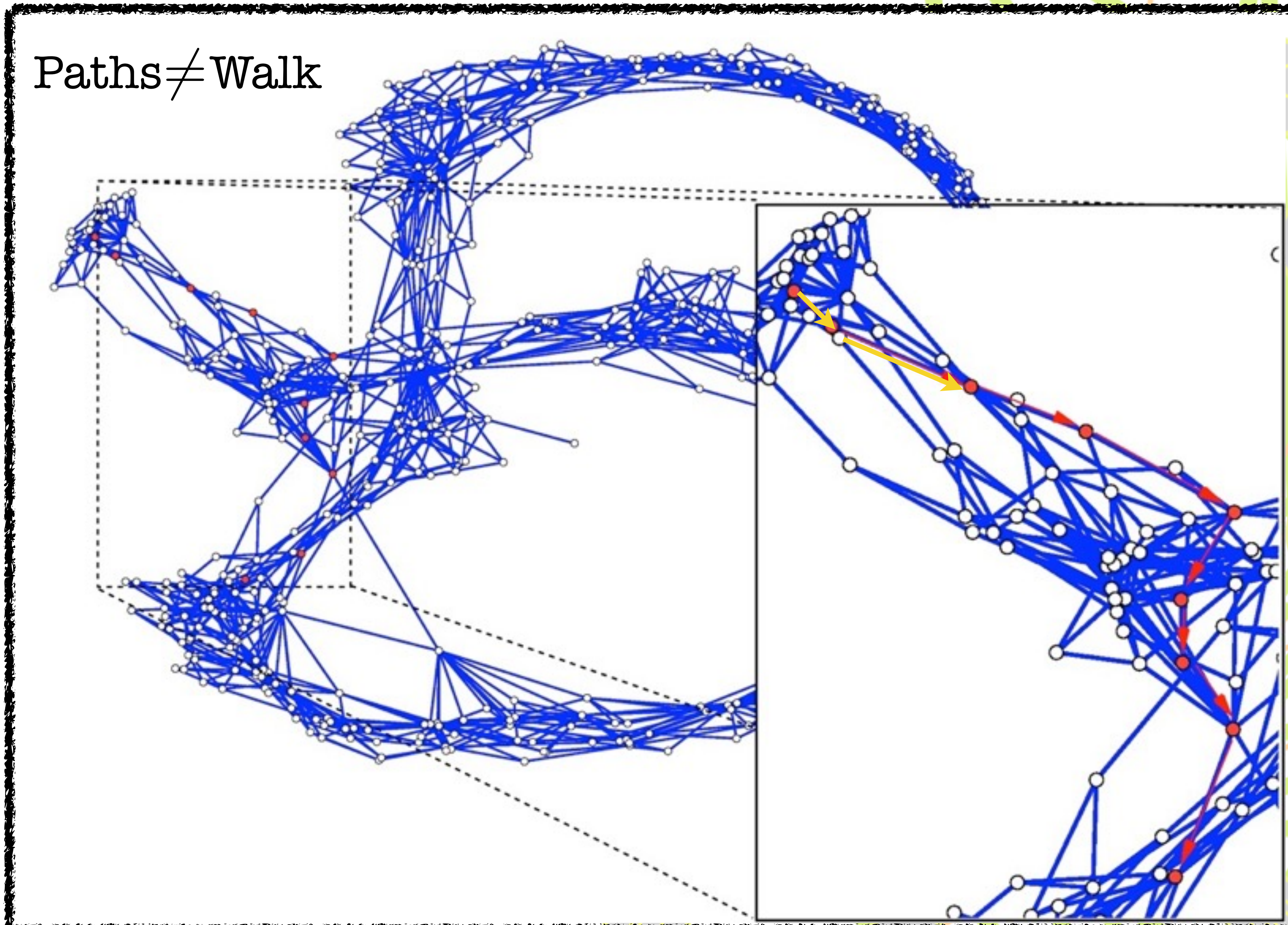


@gomezgardenes

III Distances

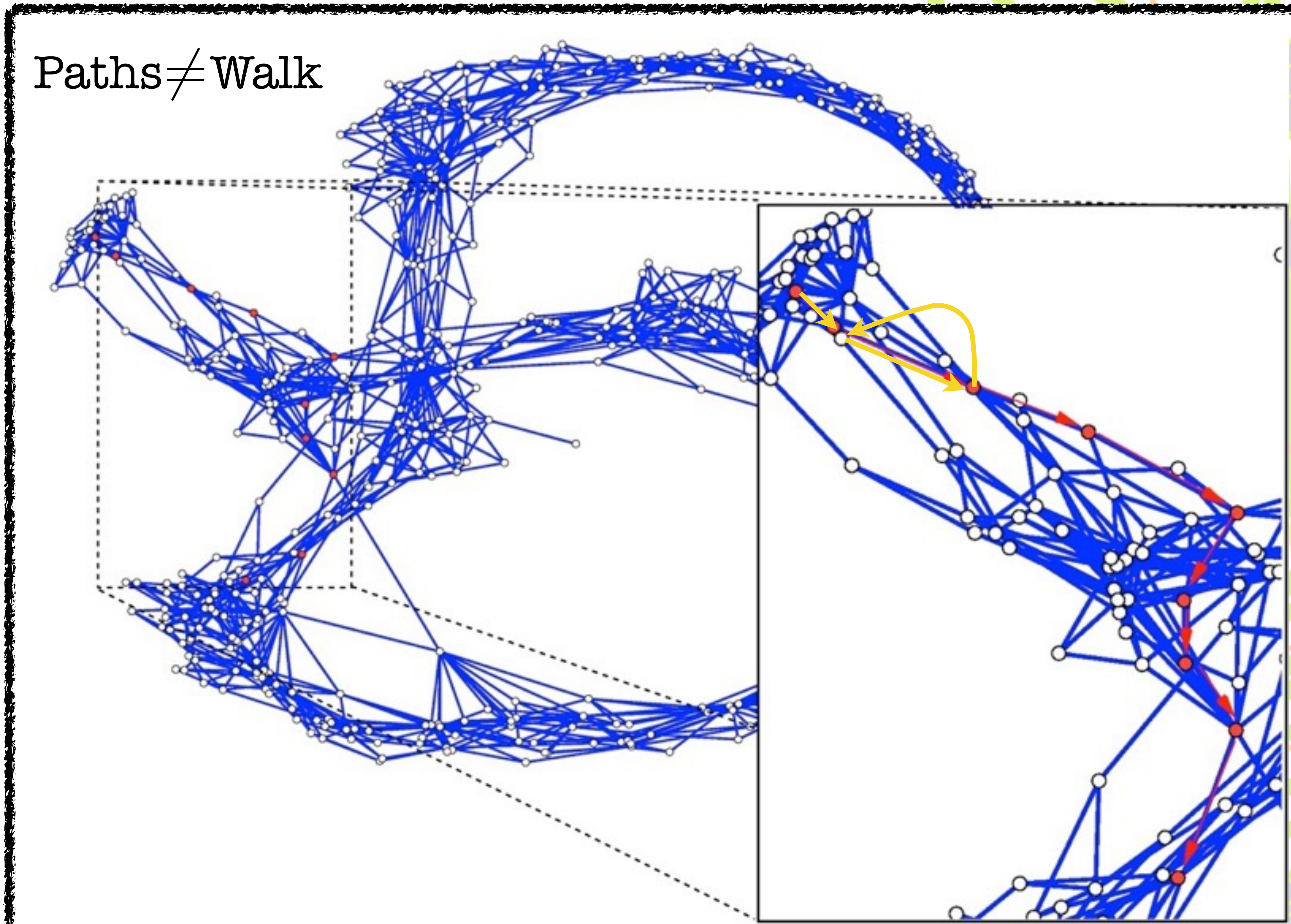


III Distances



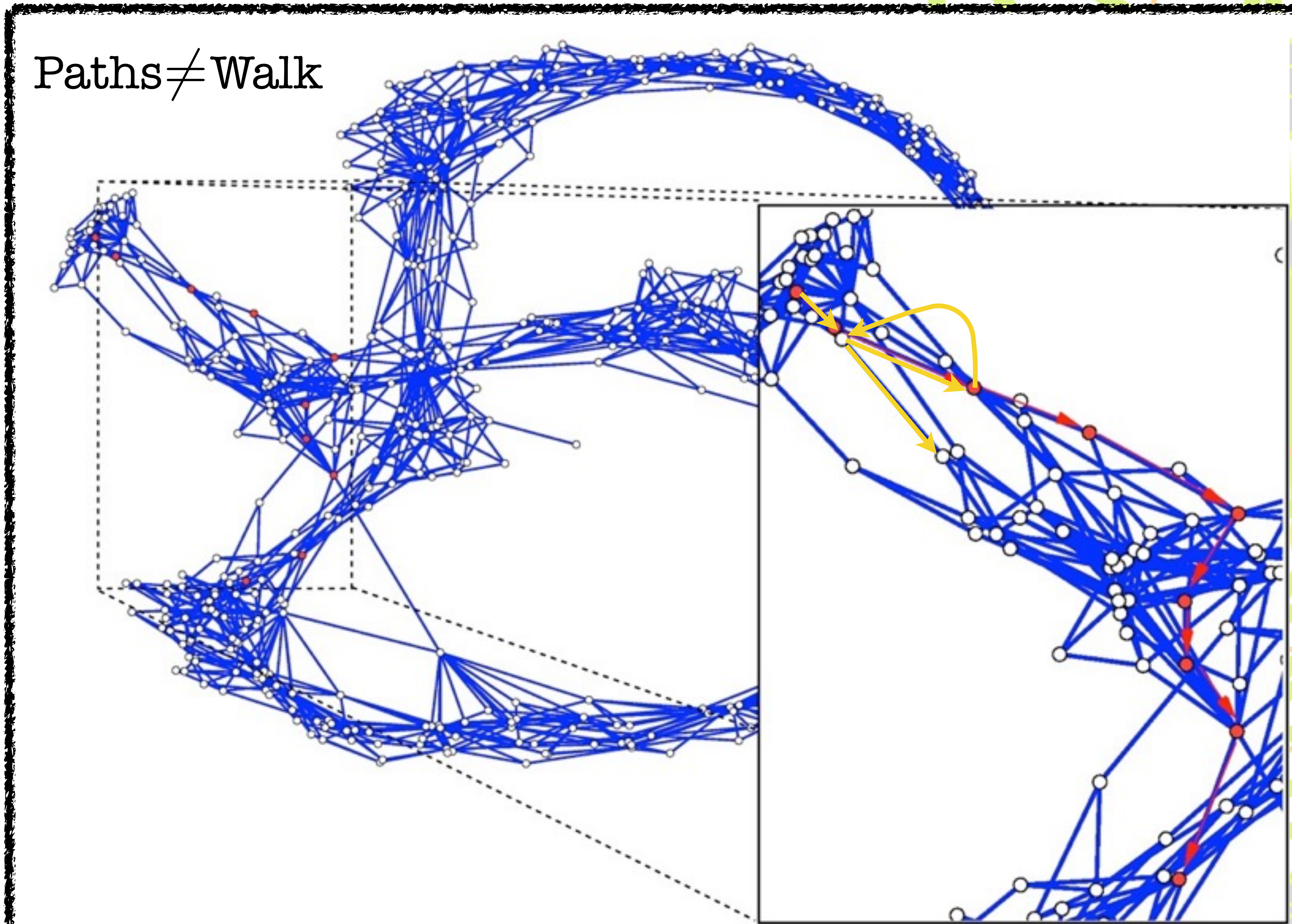
III Distances

Paths≠Walk



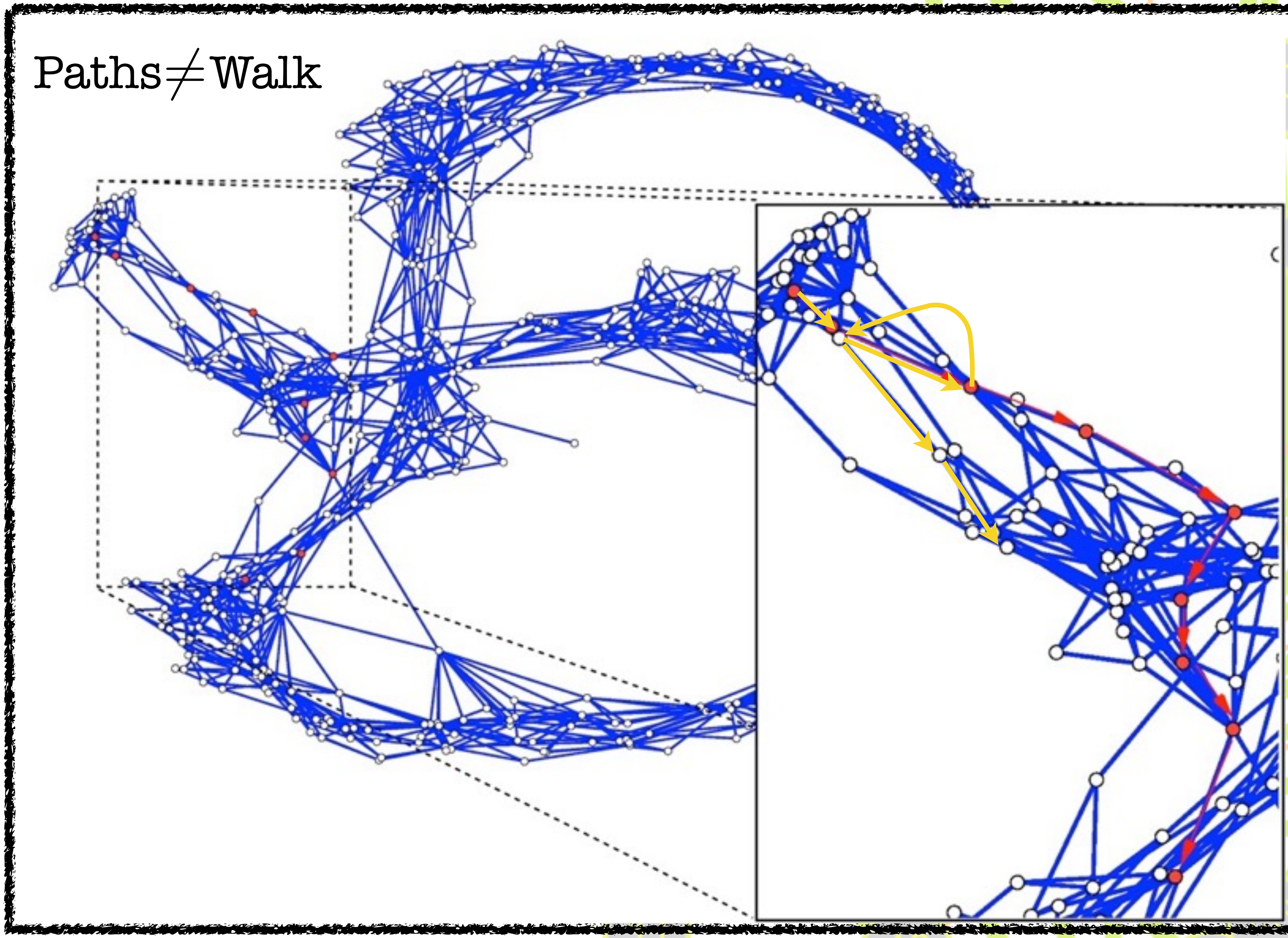
@gomezgardenes

III Distances



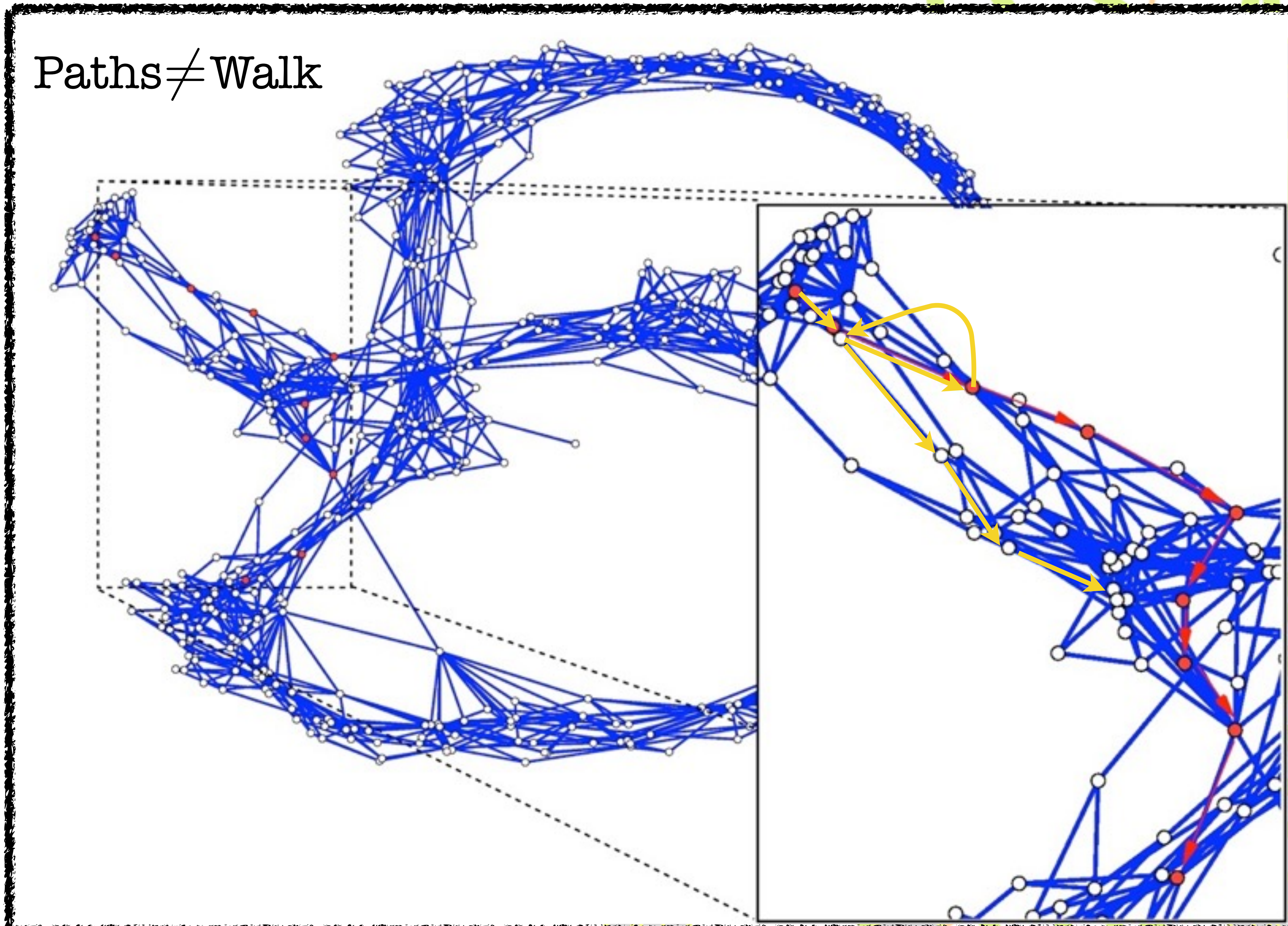
@gomezgardenes

III Distances

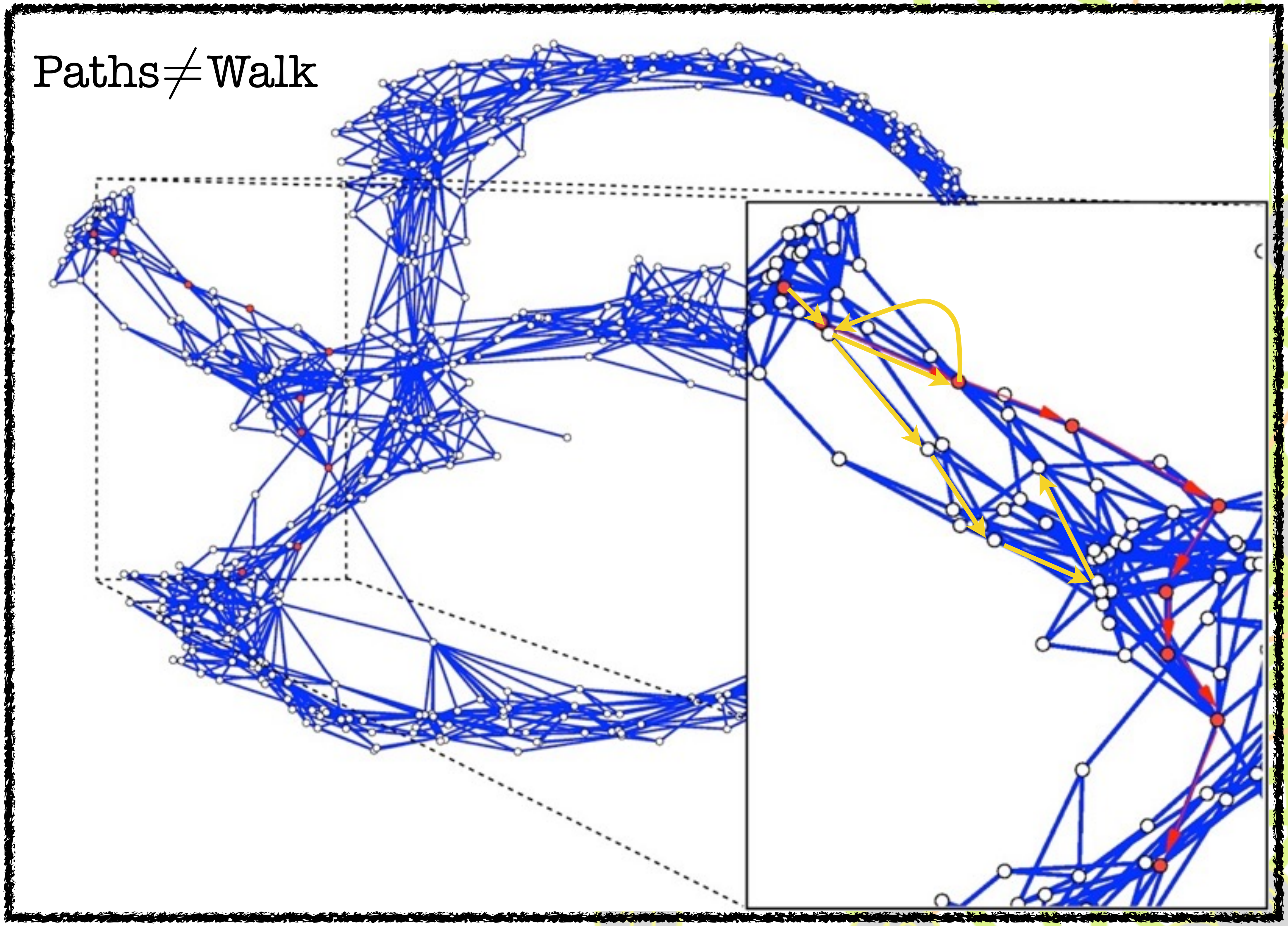


@gomezgardenes

III Distances

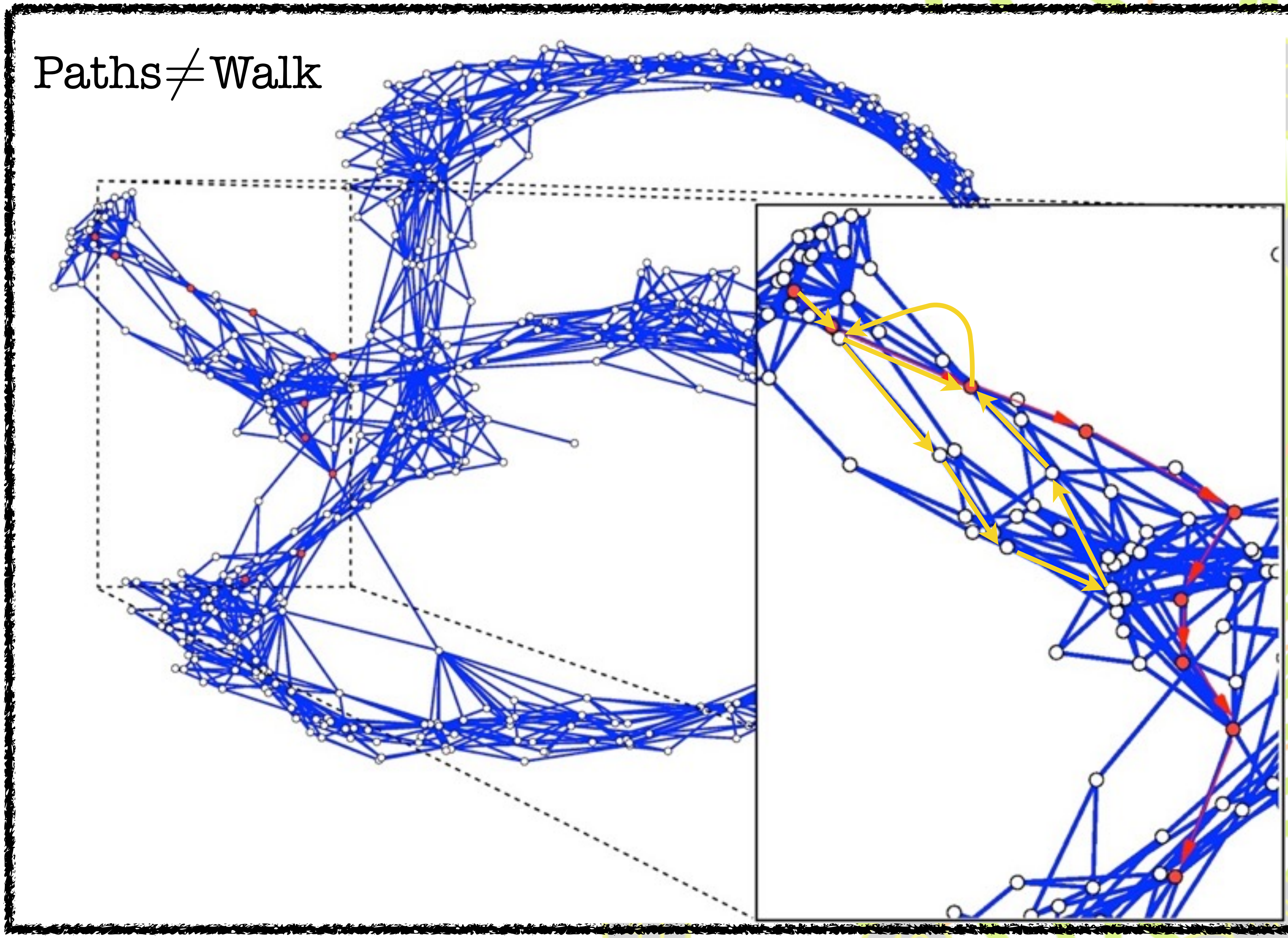


III Distances



@gomezgardenes

III Distances

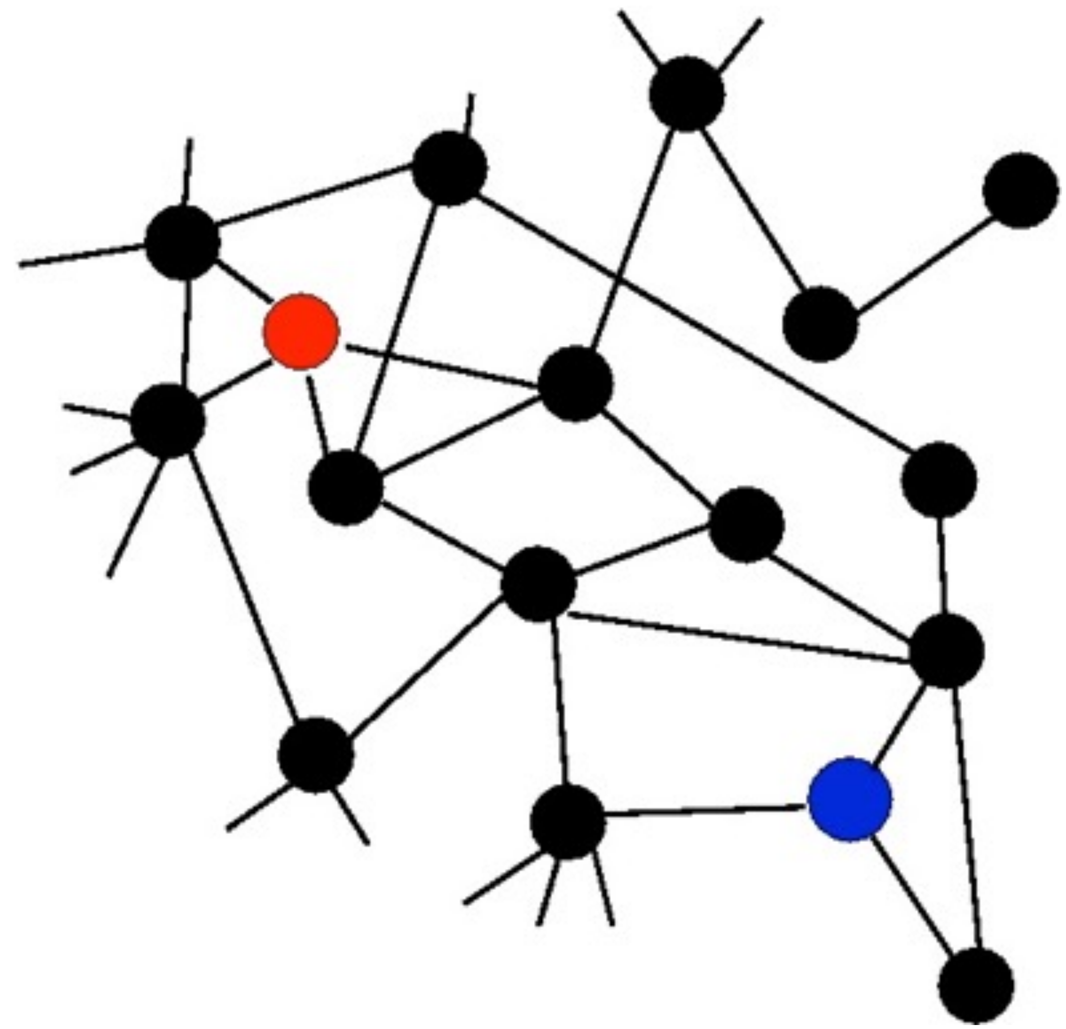


@gomezgardenes

III Distances

Distance between two nodes d_{ij} **:**

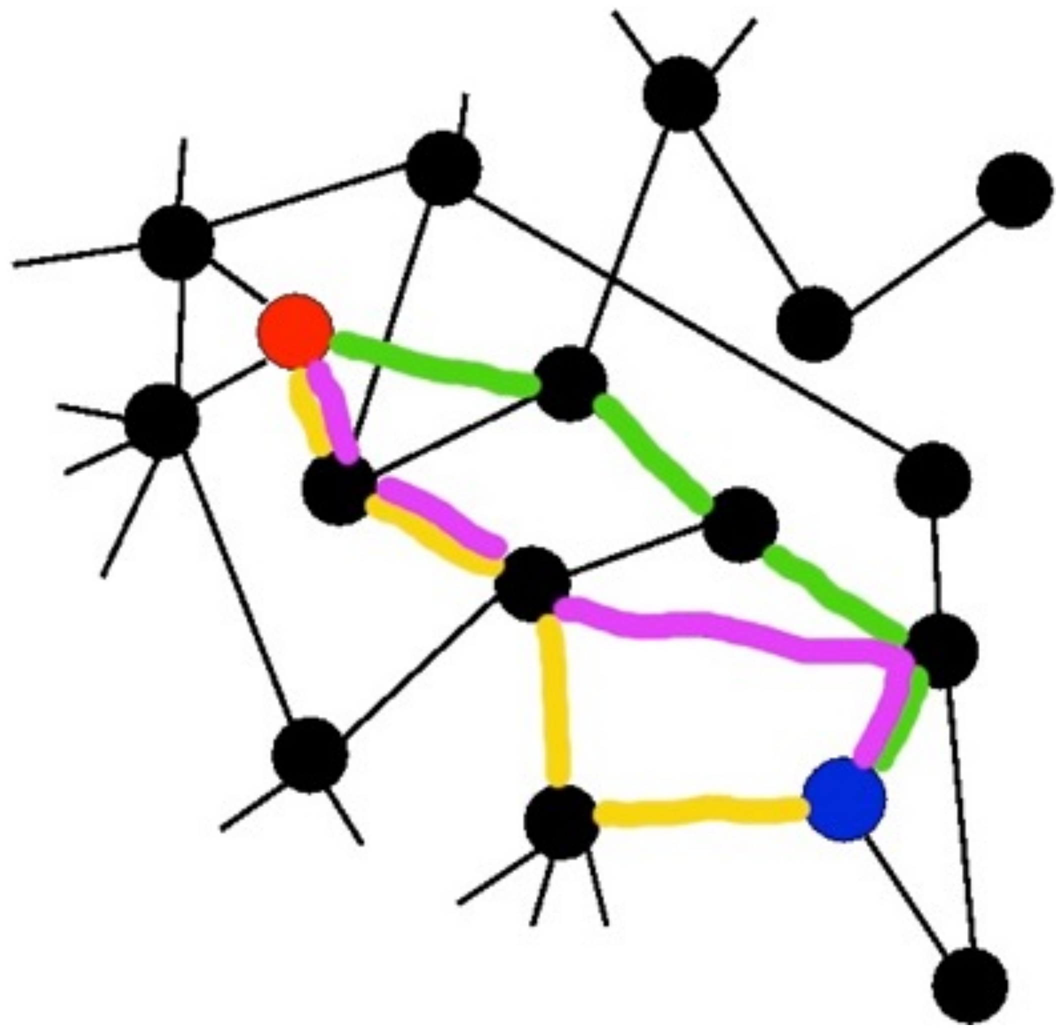
Minimum number of links to be crossed between them



III Distances

Distance between two nodes d_{ij} :

Minimum number of links to be crossed between them



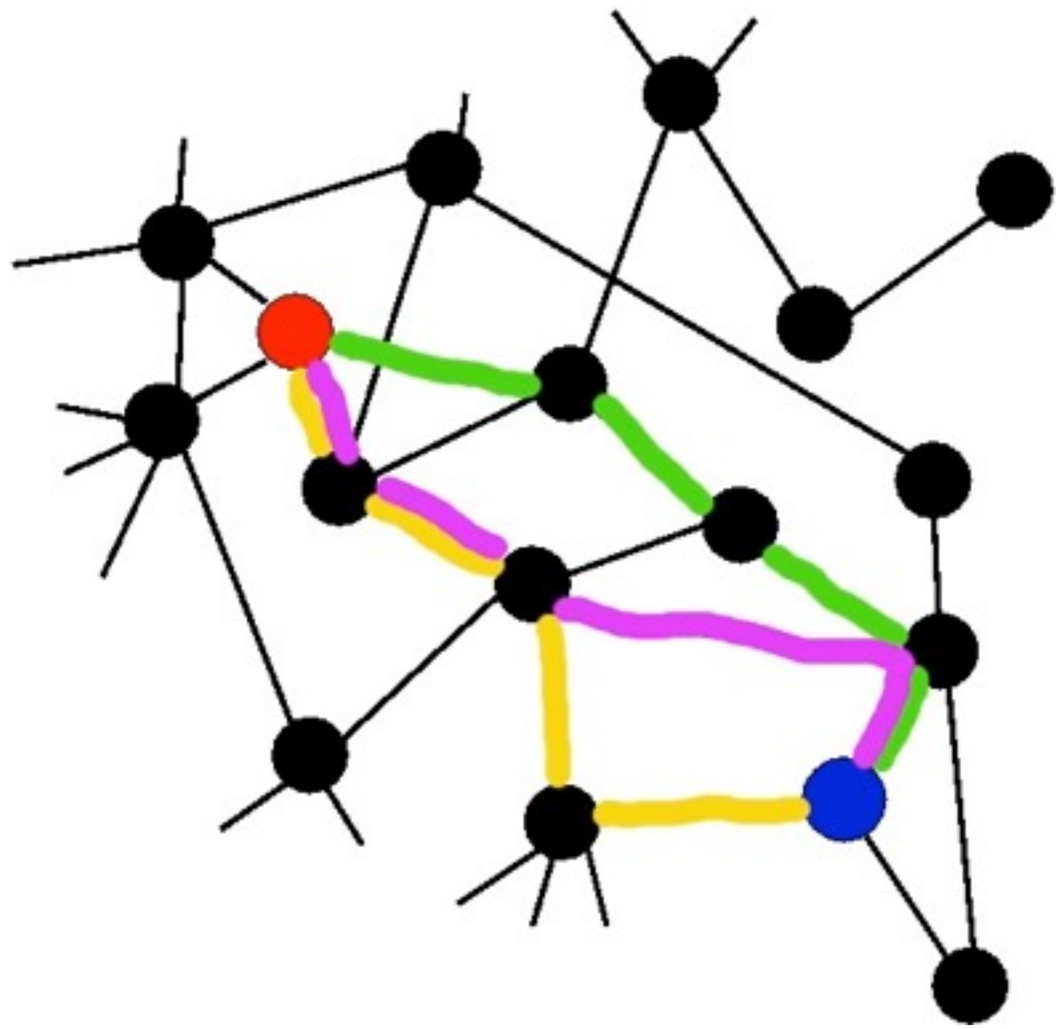
III Distances

Distance between two nodes d_{ij} :

Minimum number of links to be crossed between them

Shortest Path:

The sequence of links to be crossed



@gomezgardenes

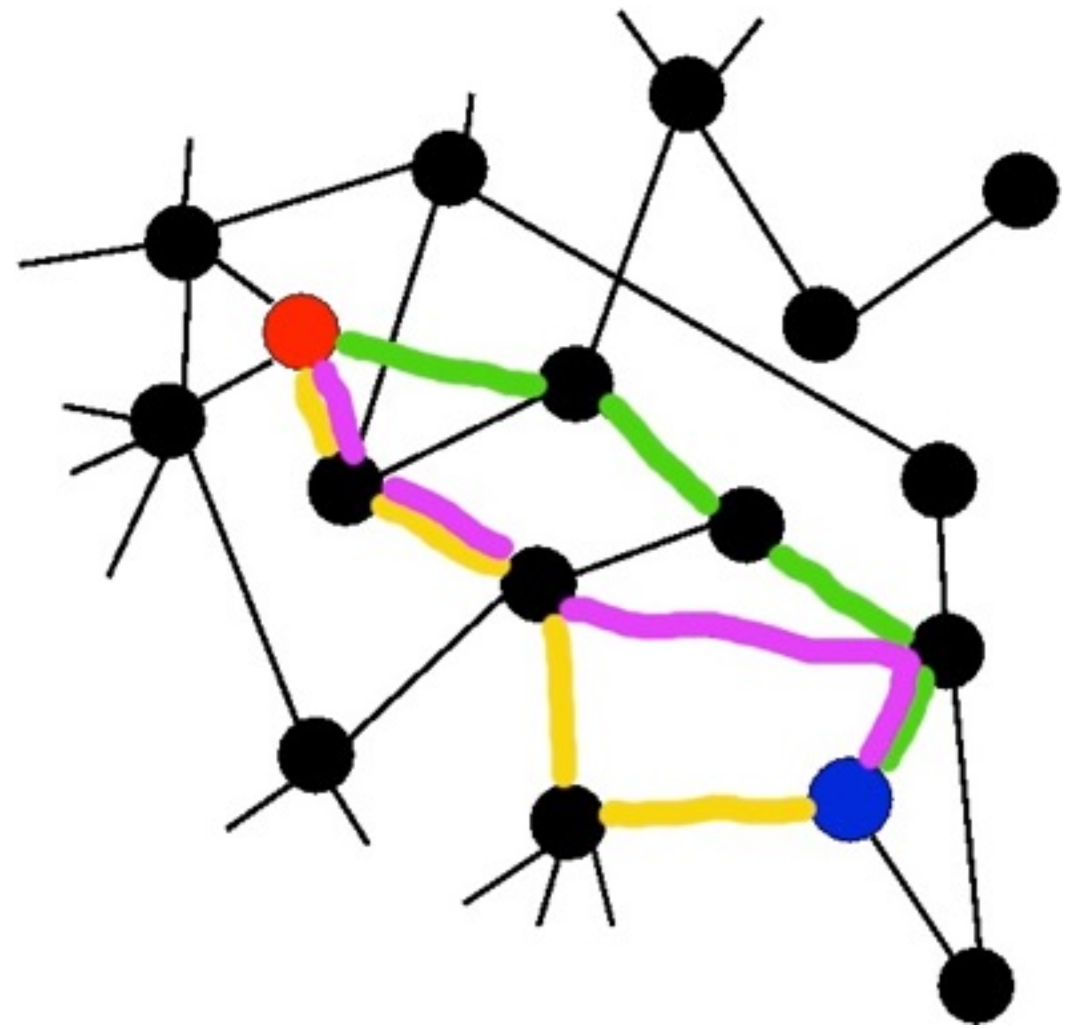
III Distances

Distance between two nodes d_{ij} :

Minimum number of links to be crossed between them

Shortest Path:

The sequence of links to be crossed



Average Path Length:

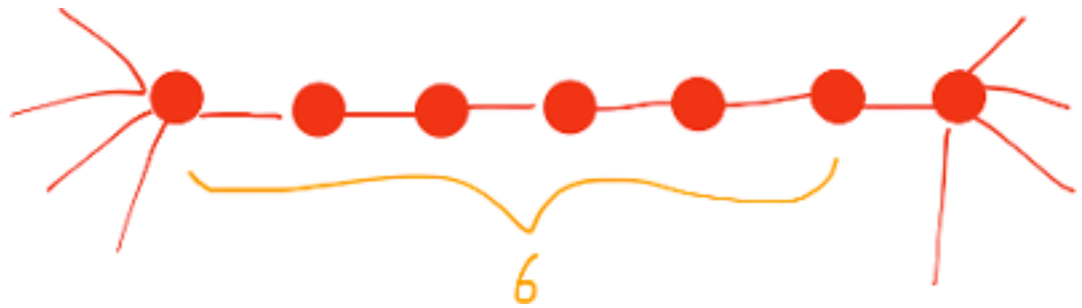
$$L = \frac{1}{N(N-1)} \sum_{i,j=1}^{N} d_{ij}$$

Diameter: $D = \max\{d_{ij}\}$

@gomezgardenes

III Distances

Small-World Phenomenon (Six Degrees of Separation)

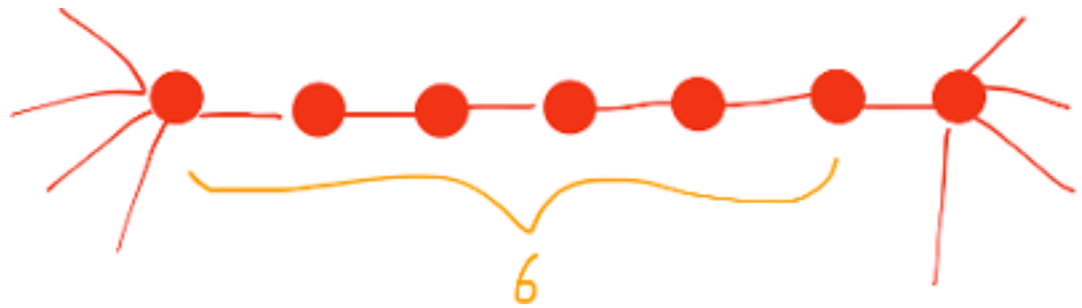


Everybody is connected to everybody else by no more than six degrees of separation
by sociologist Stanley Milgram (1967)

@gomezgardenes

III Distances

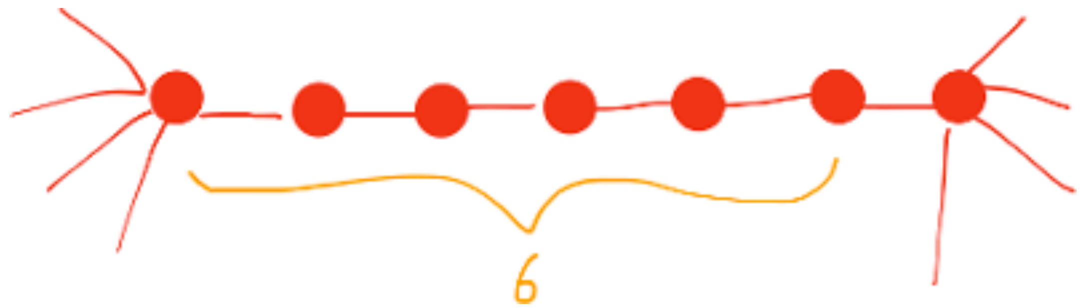
Small-World Phenomenon (Six Degrees of Separation)



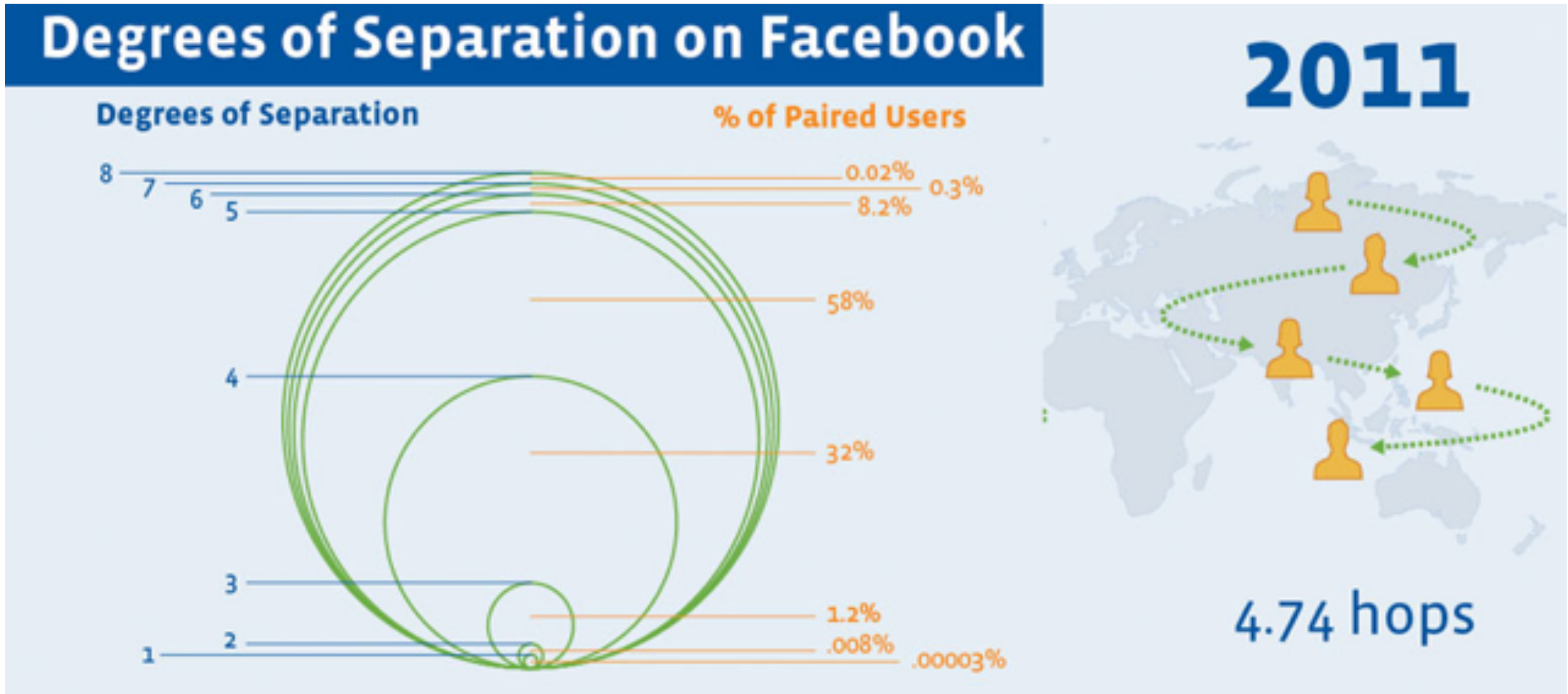
Everybody is connected to everybody else by no more than six degrees of separation
by sociologist Stanley Milgram (1967)

III Distances

Small-World Phenomenon (Six Degrees of Separation)



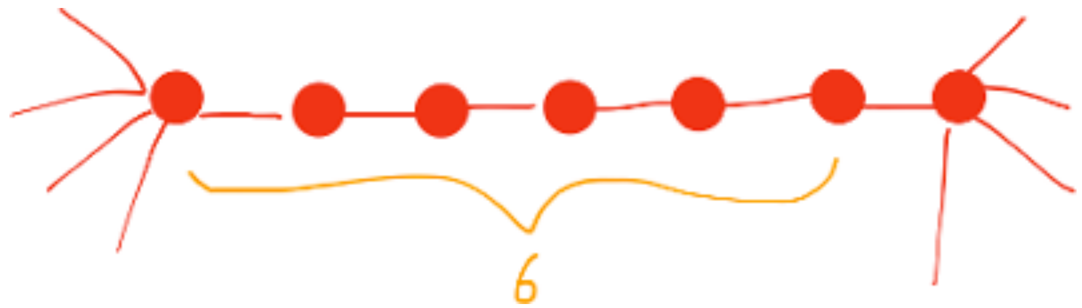
Everybody is connected to everybody else by no more than six degrees of separation
by sociologist Stanley Milgram (1967)



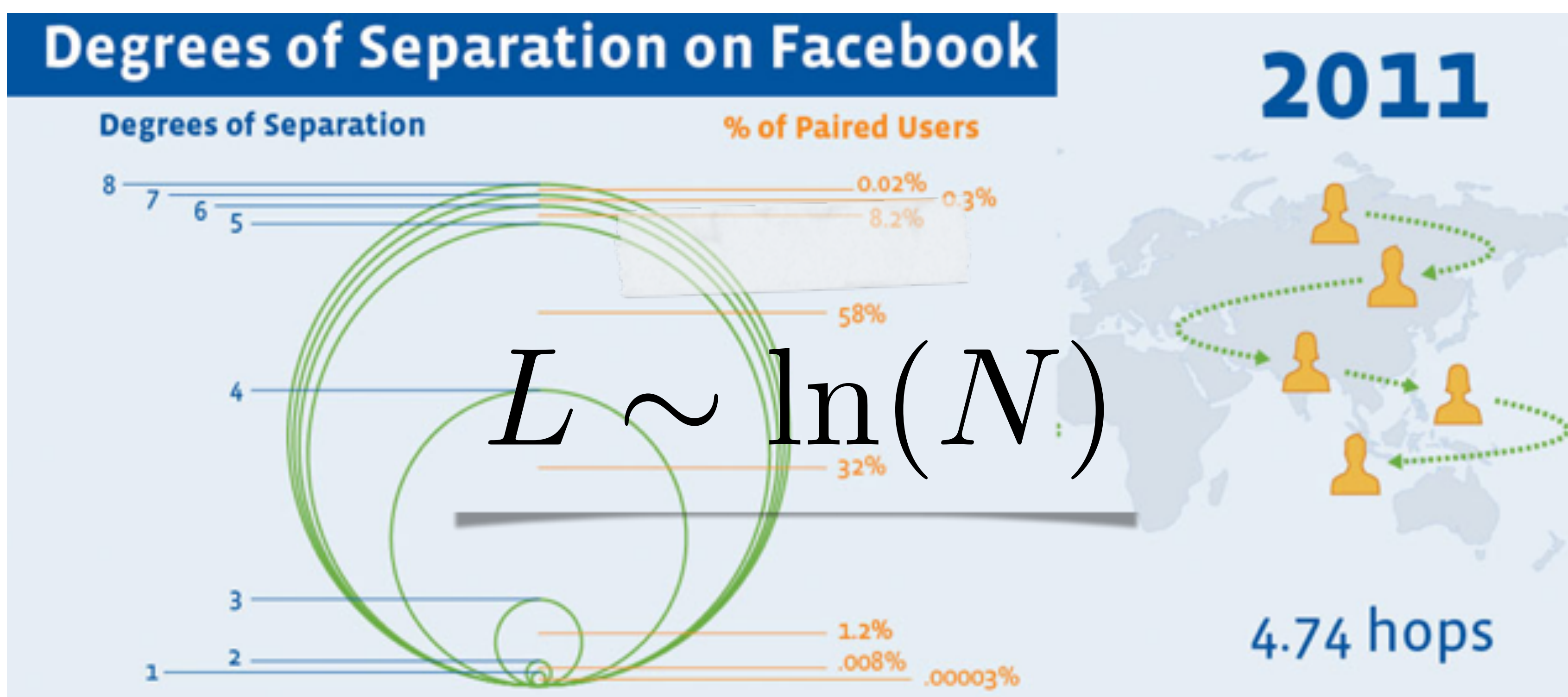
@gomezgardenes

III Distances

Small-World Phenomenon (Six Degrees of Separation)



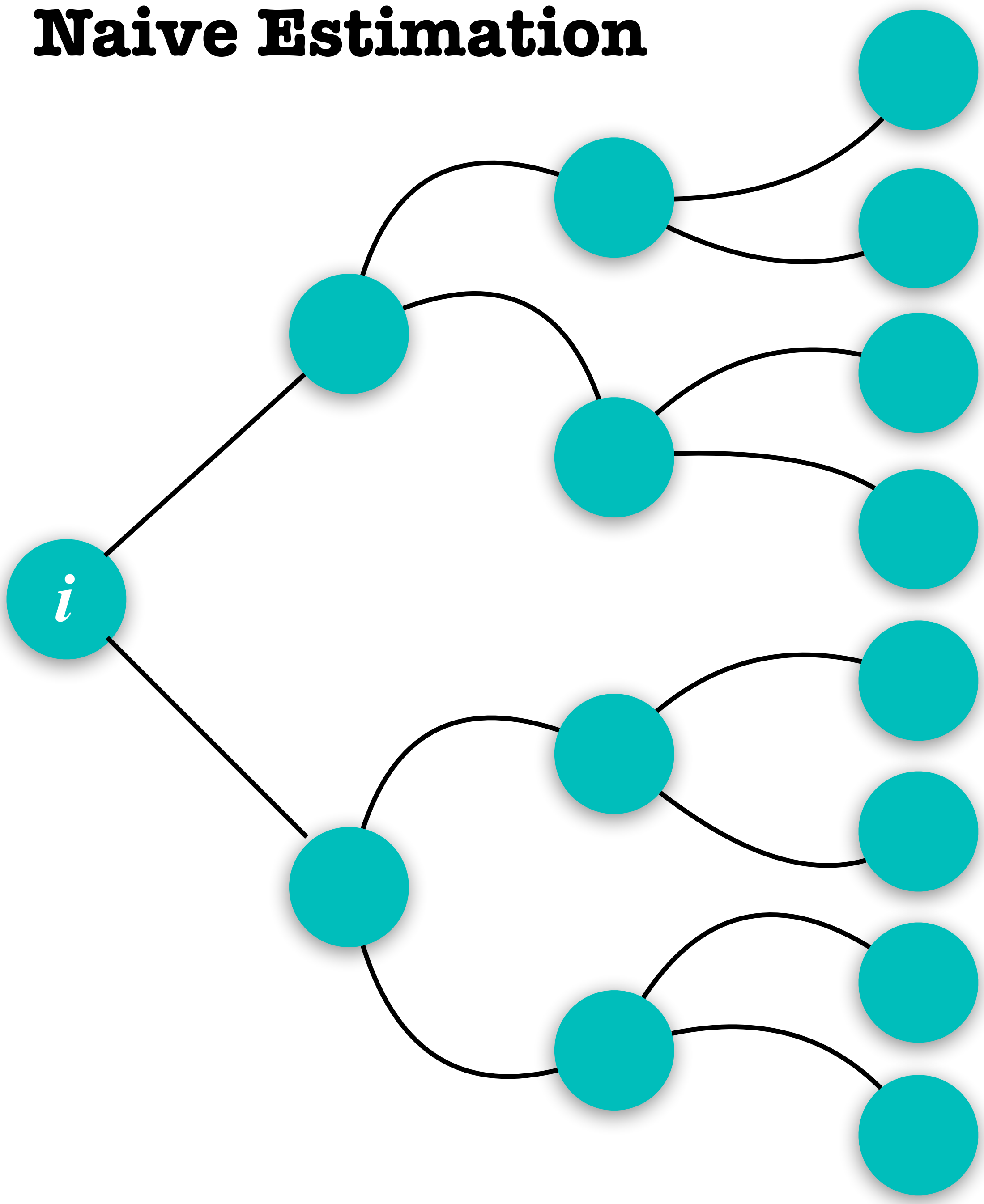
Everybody is connected to everybody else by no more than six degrees of separation
by sociologist Stanley Milgram (1967)



$$L \sim \ln(N)$$

III Distances

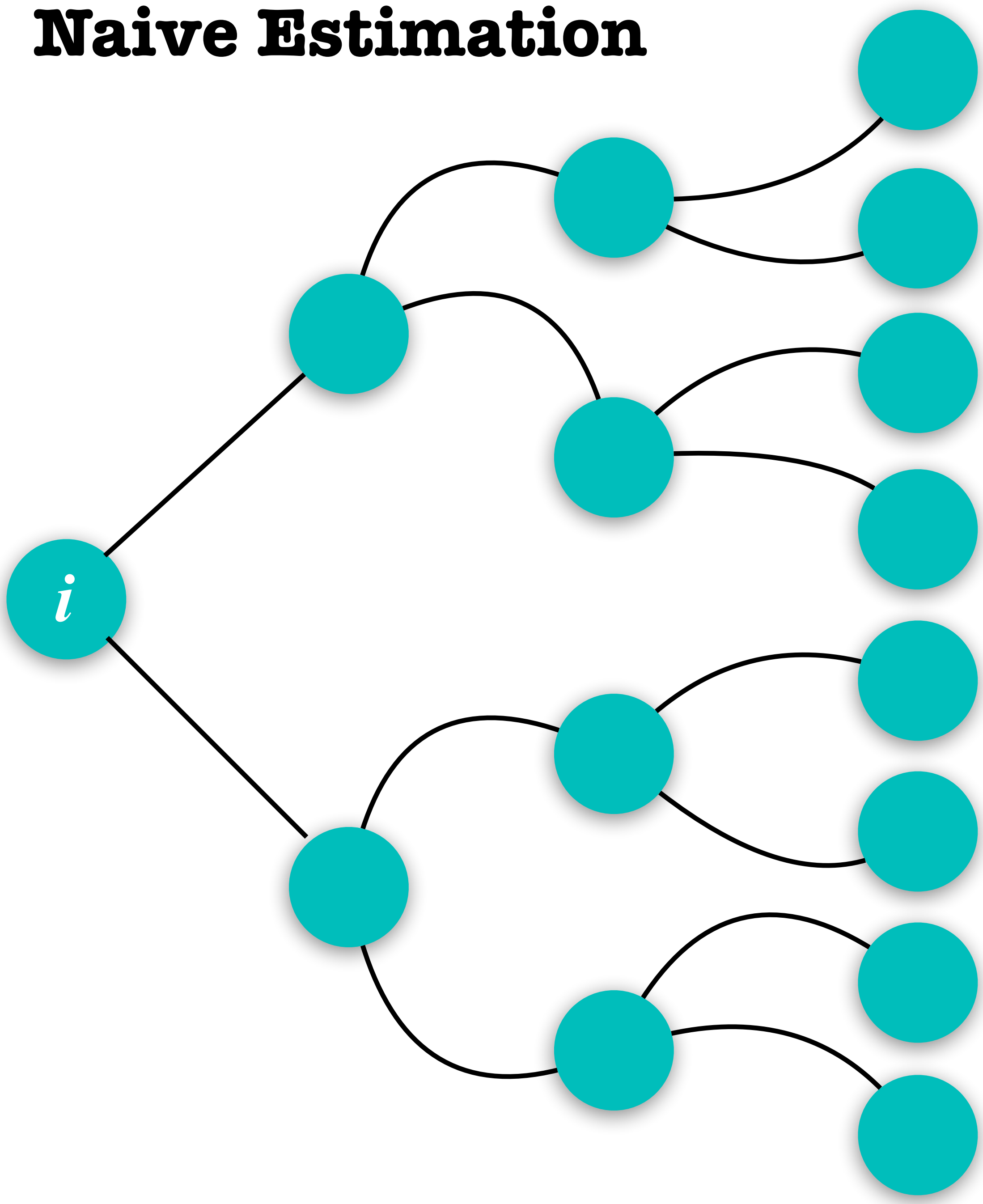
Naive Estimation



@gomezgardenes

III Distances

Naive Estimation



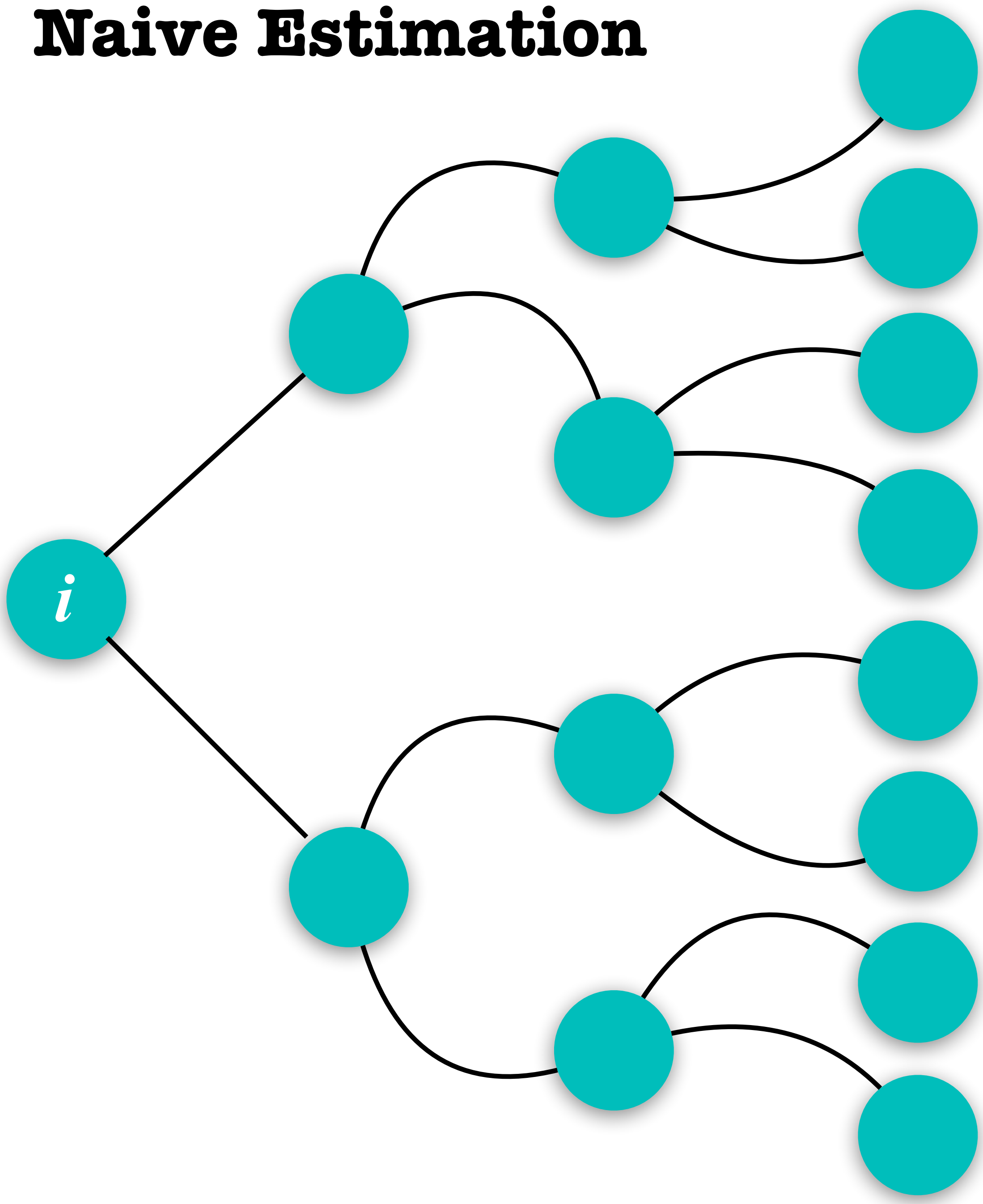
The number of nodes at distance l from node i is:

$$\langle k \rangle^l$$

@gomezgardenes

III Distances

Naive Estimation



The number of nodes at distance l from node i is:

$$\langle k \rangle^l$$

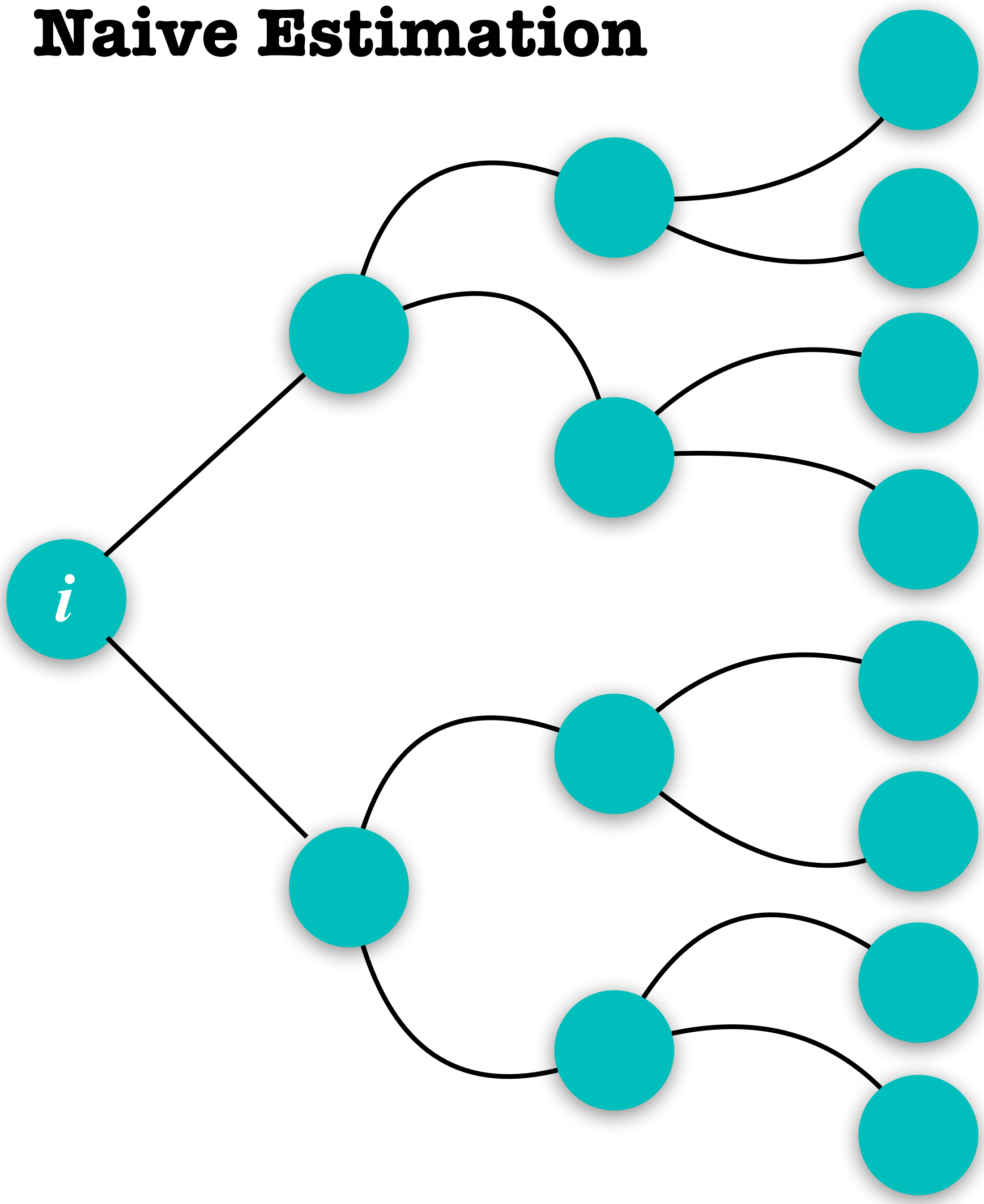
To reach all the nodes:

$$N = \sum_{l=0}^{L_{\max}} \langle k \rangle^l$$

@gomezgardenes

III Distances

Naive Estimation



The number of nodes at distance l from node i is:

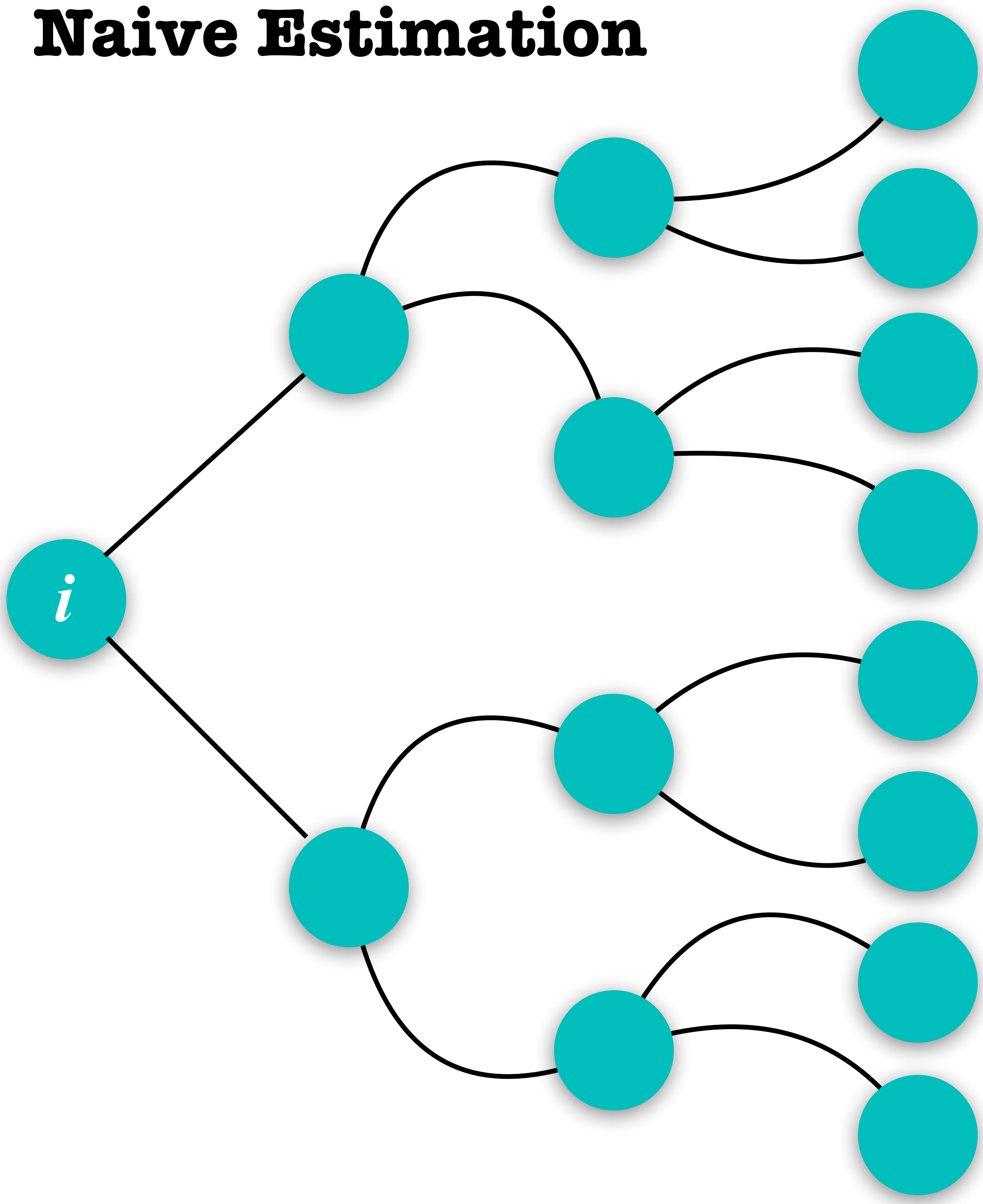
$$\langle k \rangle^l$$

To reach all the nodes:

$$N = \sum_{l=0}^{L_{\max}} \langle k \rangle^l \geq \langle k \rangle^{L_{\max}}$$

III Distances

Naive Estimation



The number of nodes at distance l from node i is:

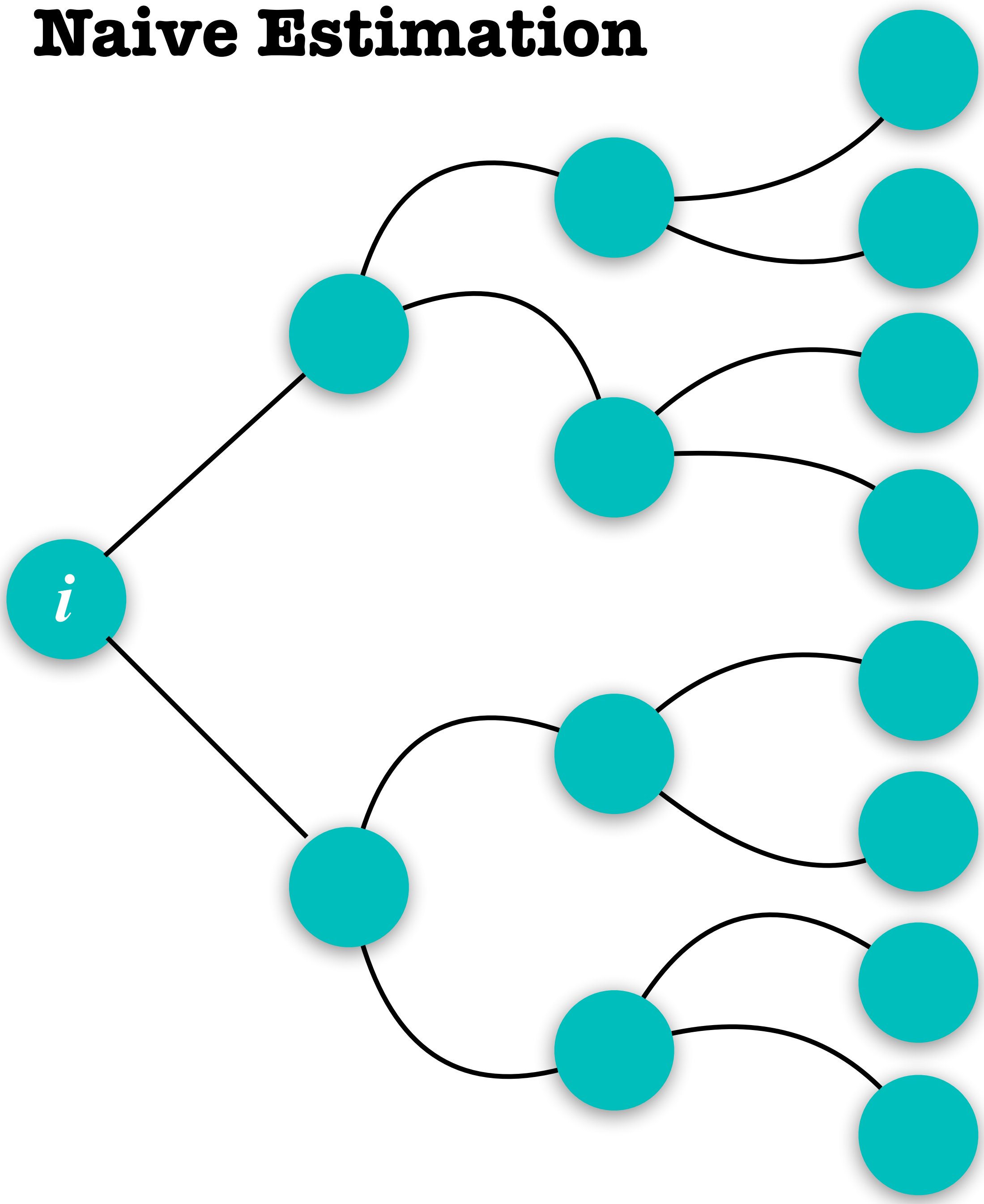
$$\langle k \rangle^l$$

To reach all the nodes:

$$N = \sum_{l=0}^{L_{\max}} \langle k \rangle^l \geq \langle k \rangle^{L_{\max}} \geq \langle k \rangle^L$$

III Distances

Naive Estimation



The number of nodes at distance l from node i is:

$$\langle k \rangle^l$$

To reach all the nodes:

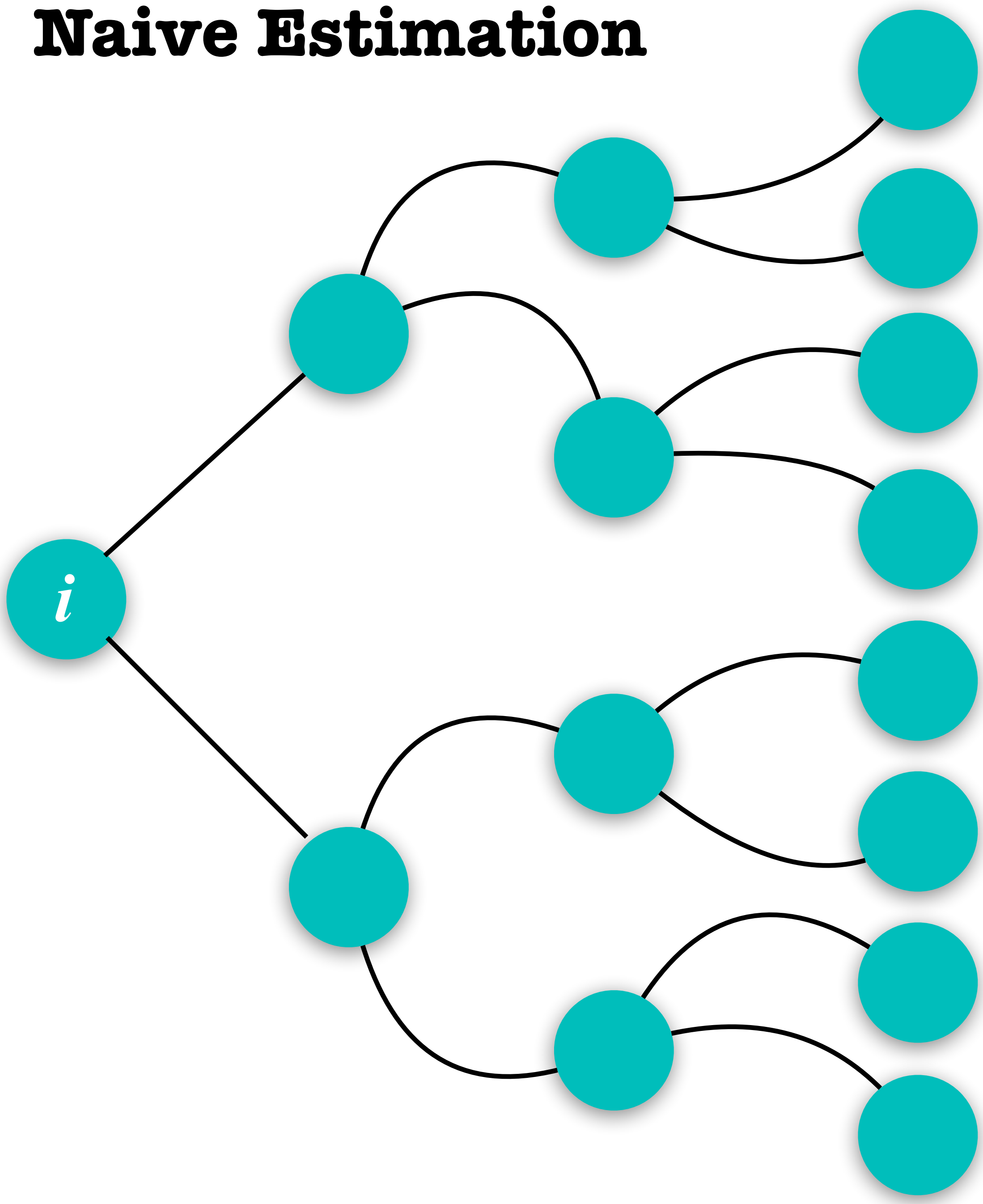
$$N = \sum_{l=0}^{L_{\max}} \langle k \rangle^l \geq \langle k \rangle^{L_{\max}} \geq \langle k \rangle^L$$

$$\log(N) \geq L \log(\langle k \rangle)$$

@gomezgardenes

III Distances

Naive Estimation



The number of nodes at distance l from node i is:

$$\langle k \rangle^{l}$$

To reach all the nodes:

$$N = \sum_{l=0}^{L_{\max}} \langle k \rangle^{l} \geq \langle k \rangle^{L_{\max}} \geq \langle k \rangle^{L}$$

$$\log(N) \geq L \log(\langle k \rangle)$$

$$L \leq \frac{\log(N)}{\log(\langle k \rangle)}$$

IV Correlations

$P(k', k)$: Probability that two nodes of degree k and k' are linked

Detailed balance Equation for Networks

$$P(k', k) = kP(k)P(k'|k) = k'P(k')P(k|k')$$

IV Correlations

$P(k',k)$: Probability that two nodes of degree k and k' are linked

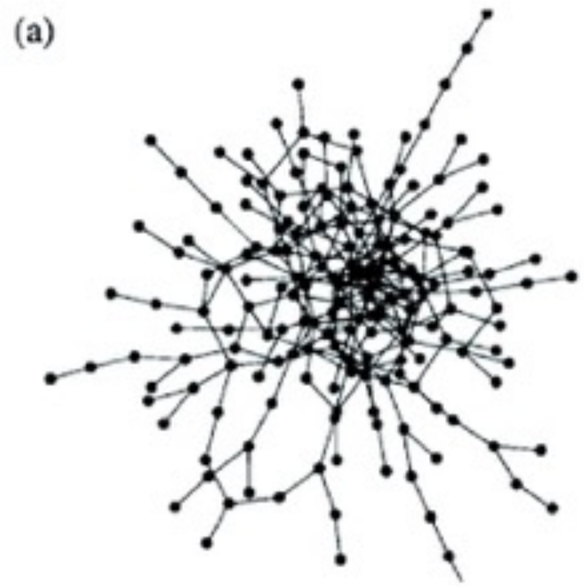
Detailed balance Equation for Networks

$$P(k',k) = kP(k)P(k'|k) = k'P(k')P(k|k')$$

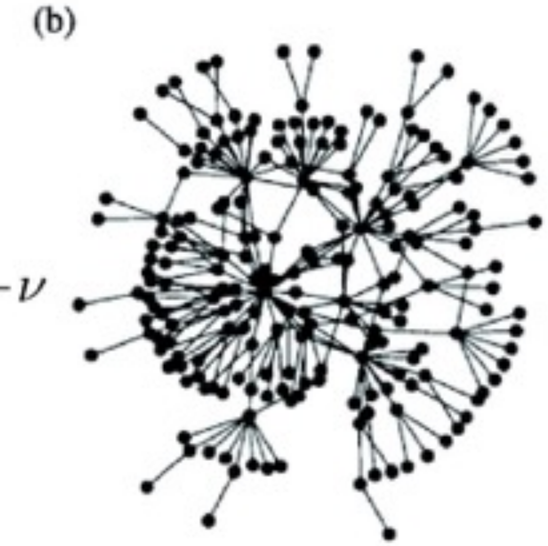
Two ways of measuring

$$k_{nn} = \sum_{k'} k'P(k'|k) = f(k) \qquad r = \frac{\langle k_i k_j \rangle - \langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle^2}$$

IV Correlations



$$k_{nn}(k) \sim k^{-\nu}$$



Positive correlations
ASSORTATIVE NETWORKS

$\nu < 0$

$r > 0$

Negative correlations
DISASSORTATIVE NETWORKS

$\nu > 0$

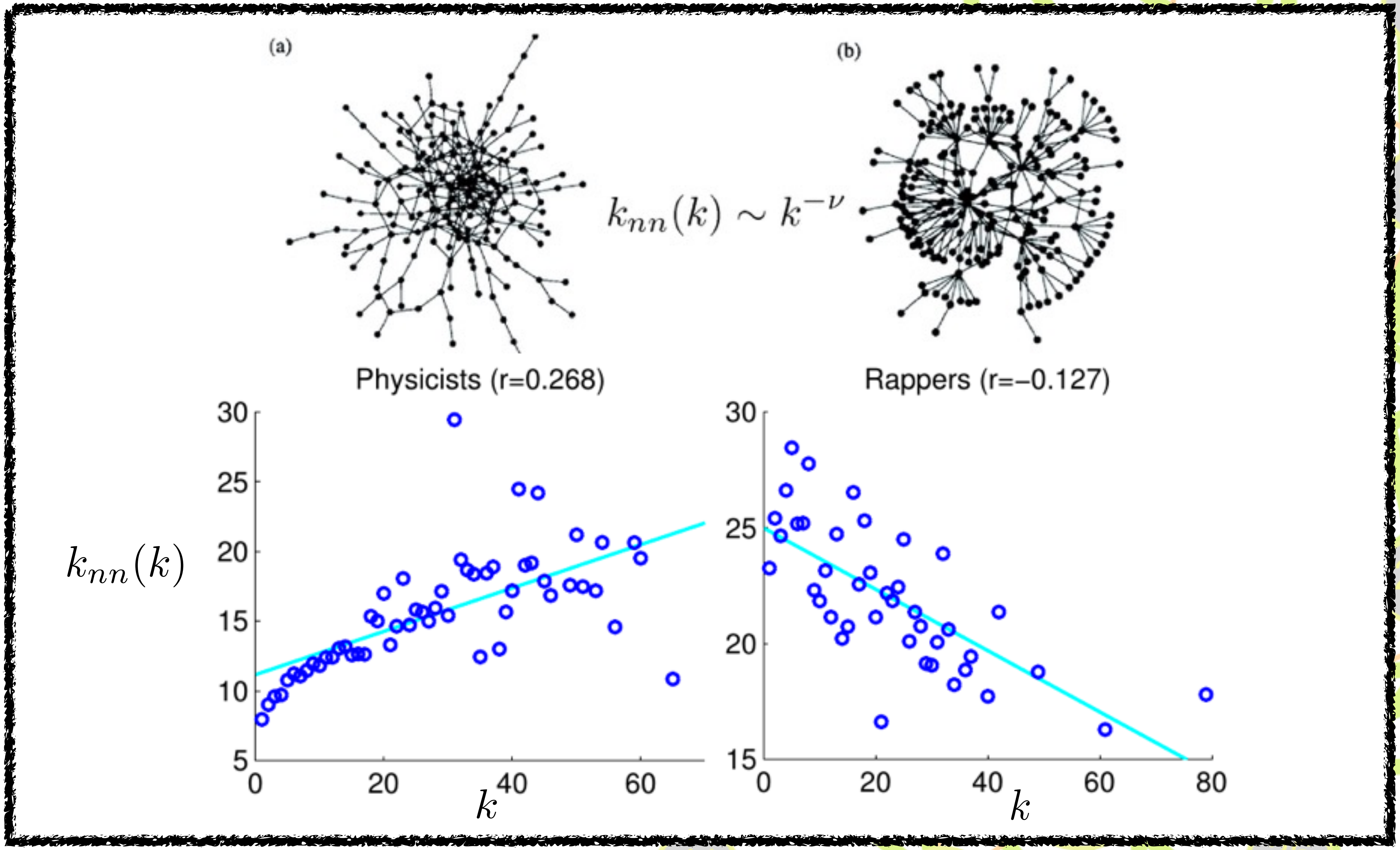
$r < 0$

$$k_{nn} = \sum_{k'} k' P(k'|k) = f(k)$$

$$r = \frac{\langle k_i k_j \rangle - \langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle^2}$$

@gomezgardenes

IV Correlations



@gomezgardenes

Networks' Taxonomy

	Network	Type	Nodes	Links	<k>	L		Clustering		Corr. (r)
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1			
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16	
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029
	sexual contacts	undirected	2 810				3.2			
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7			
	citation network	directed	783 339	6 716 198	8.57		3.0/–			
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44	
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263
	freshwater food web	directed	92	997	10.84	1.90	–	0.40	0.48	−0.326
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226

NEWMAN, SIAM REVIEWS 45, 167 (2003)

@gomezgardenes

Networks' Taxonomy

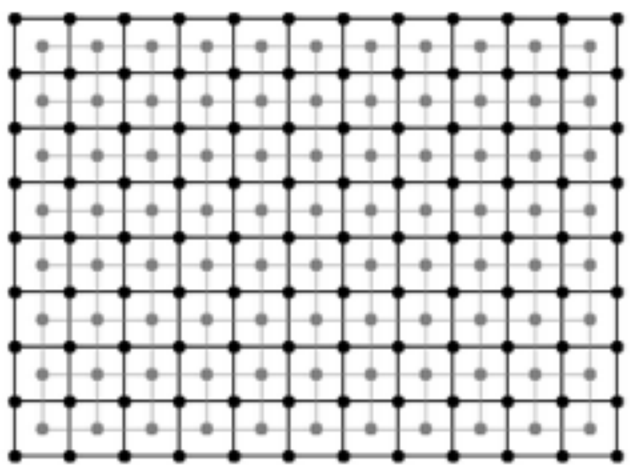
	Network	Type	Nodes	Links	<k>	L		Clustering		Corr. (r)
Social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1			
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16	
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029
	sexual contacts	undirected	2 810				3.2			
Information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7			
	citation network	directed	783 339	6 716 198	8.57		3.0/–			
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44	
Technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366
Biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263
	freshwater food web	directed	92	997	10.84	1.90	–	0.40	0.48	−0.326
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226

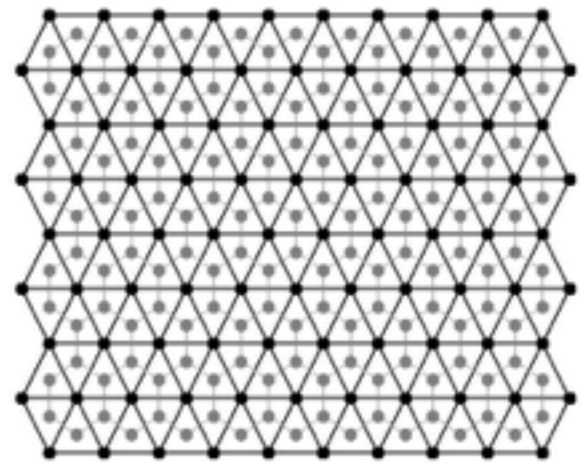
NEWMAN, SIAM REVIEWS 45, 167 (2003)

@gomezgardenes

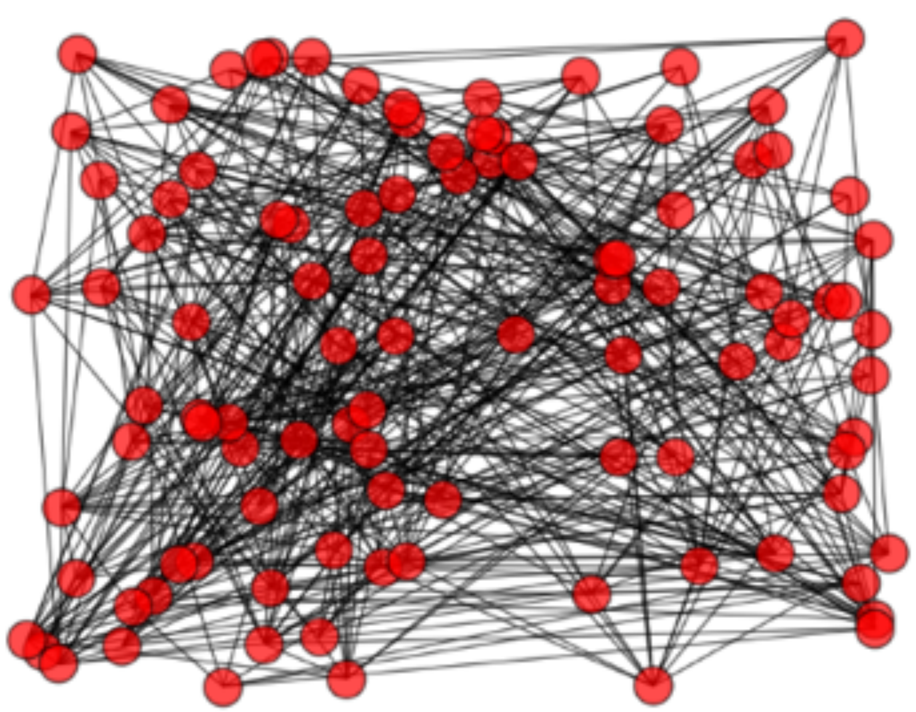
Why Complex?

- Not regular/ordered





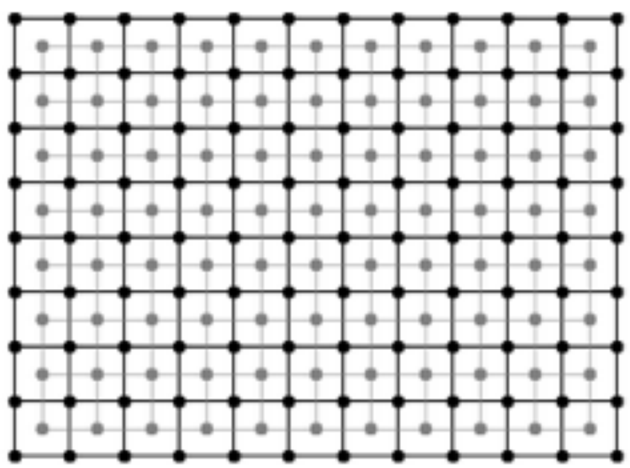
- Not completely random

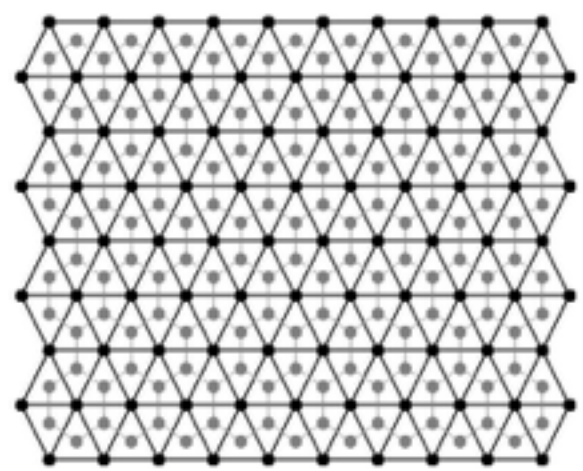


@gomezgardenes

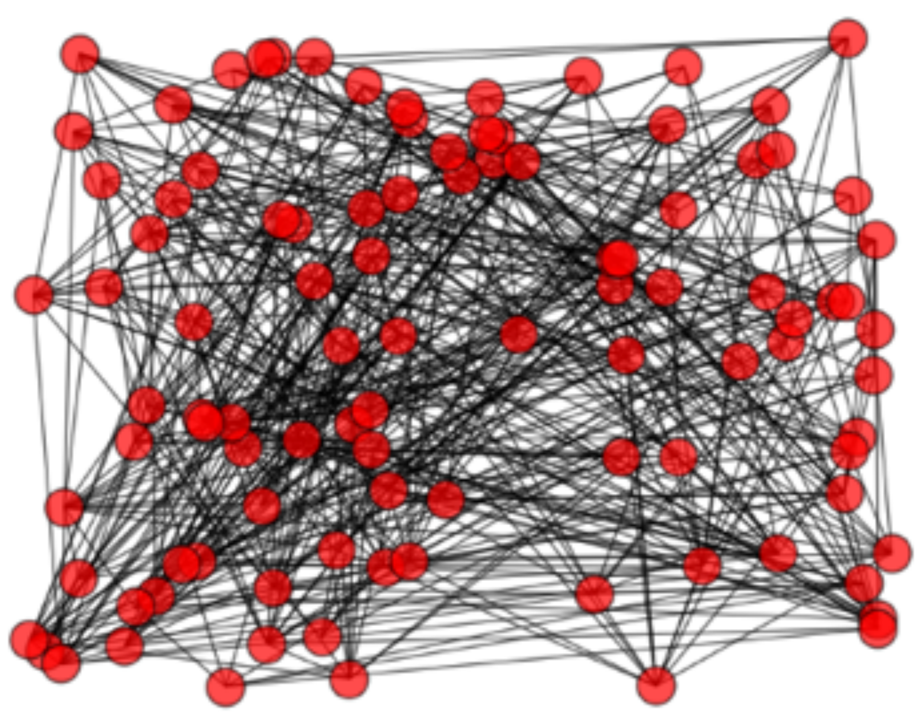
Why Complex?

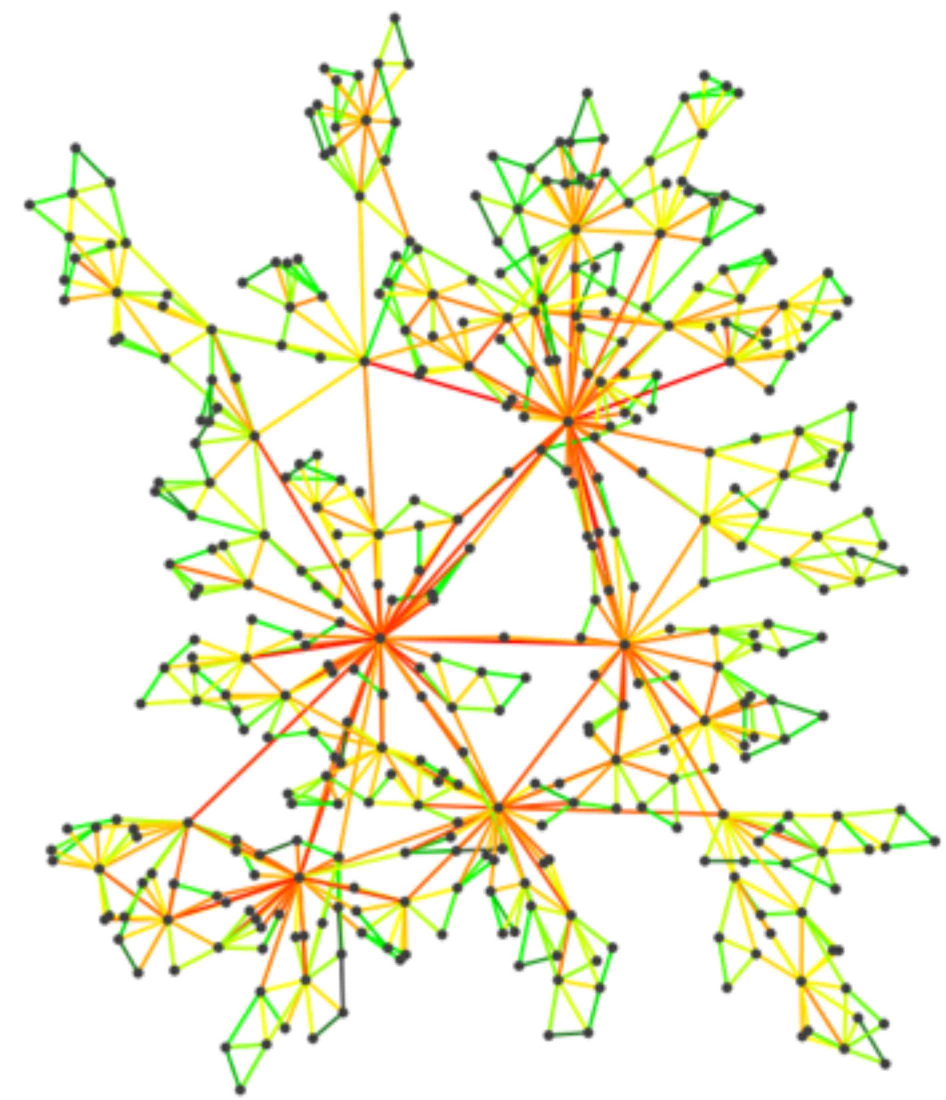
- Not regular/ordered





- Not completely random





@gomezgardenes

What are those important actors?

“Two sides of the same node”

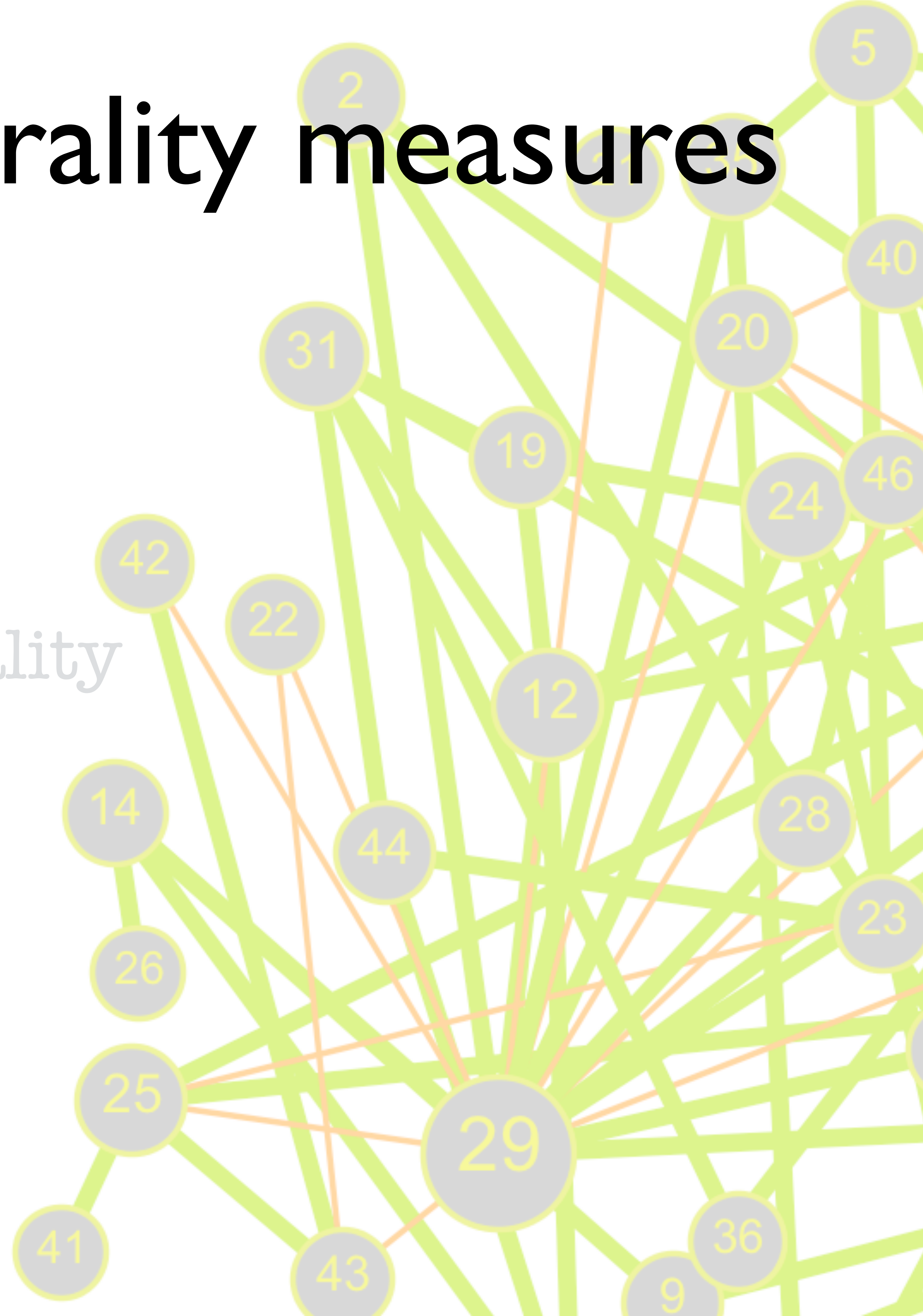
VS

Ideal targets for attacks

Good places to allocate controllers
Good candidates for being vaccinated

Main Centrality measures

- Degree
- Eigenvector Centrality
- Closeness
- Betweenness
- ...



Main Centrality measures

- Degree
- Eigenvector Centrality
- Closeness
- Betweenness
- ...

@gomezgardenes

I Eigenvector Centrality

Centrality measure of a node that takes (also) into account the importance of its neighbors

Recursive definition:

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} x_j^*$$

P. BONACICH, JOURNAL OF MATHEMATICAL SOCIOLOGY 2, 113 (1972)

I Eigenvector Centrality

Centrality measure of a node that takes (also) into account the importance of its neighbors

Recursive definition:

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} x_j^*$$

$$\frac{1}{\alpha}\mathbf{x}^* = A\mathbf{x}^*$$

High value when being important and/or
connected to important (high degree) nodes

P. BONACICH, JOURNAL OF MATHEMATICAL SOCIOLOGY 2, 113 (1972)

I Eigenvector Centrality

Centrality measure of a node that takes (also) into account the importance of its neighbors

Recursive definition:

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} x_j^*$$

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} \frac{x_j^*}{k_j} + \frac{(1-\alpha)}{N}$$

$$\frac{1}{\alpha}\mathbf{x}^* = A\mathbf{x}^*$$

High value when being important and/or connected to important (high degree) nodes

P. BONACICH, JOURNAL OF MATHEMATICAL SOCIOLOGY 2, 113 (1972)

@gomezgardenes

I Eigenvector Centrality

Centrality measure of a node that takes (also) into account the importance of its neighbors

Recursive definition:

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} x_j^*$$

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} \frac{x_j^*}{k_j} + \frac{(1-\alpha)}{N}$$

Pagerank centrality

$$\frac{1}{\alpha}\mathbf{x}^* = A\mathbf{x}^*$$

High value when being important and/or connected to important (high degree) nodes

P. BONACICH, JOURNAL OF MATHEMATICAL SOCIOLOGY 2, 113 (1972)

@gomezgardenes

I Eigenvector Centrality

Centrality measure of a node that takes (also) into account the importance of its neighbors

Recursive definition:

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} x_j^*$$

$$x_i^* = \alpha \sum_{j=1}^{N} A_{ij} \frac{x_j^*}{k_j} + \frac{(1-\alpha)}{N}$$

Pagerank centrality

$$\frac{1}{\alpha}\mathbf{x}^* = A\mathbf{x}^*$$

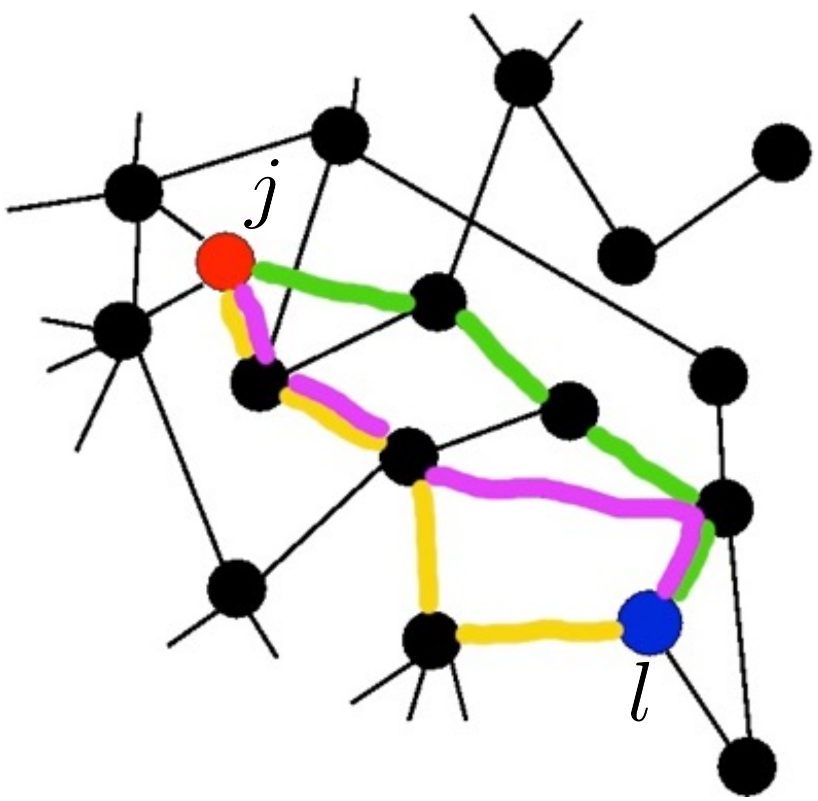
High value when being important and/or connected to important (high degree) nodes

P. BONACICH, JOURNAL OF MATHEMATICAL SOCIOLOGY 2, 113 (1972)

II Closeness

Considers the distances between a node and the rest of the network

$$c_j = \frac{1}{N^{-1}\sum_{l=1}^{N} d_{jl}}$$

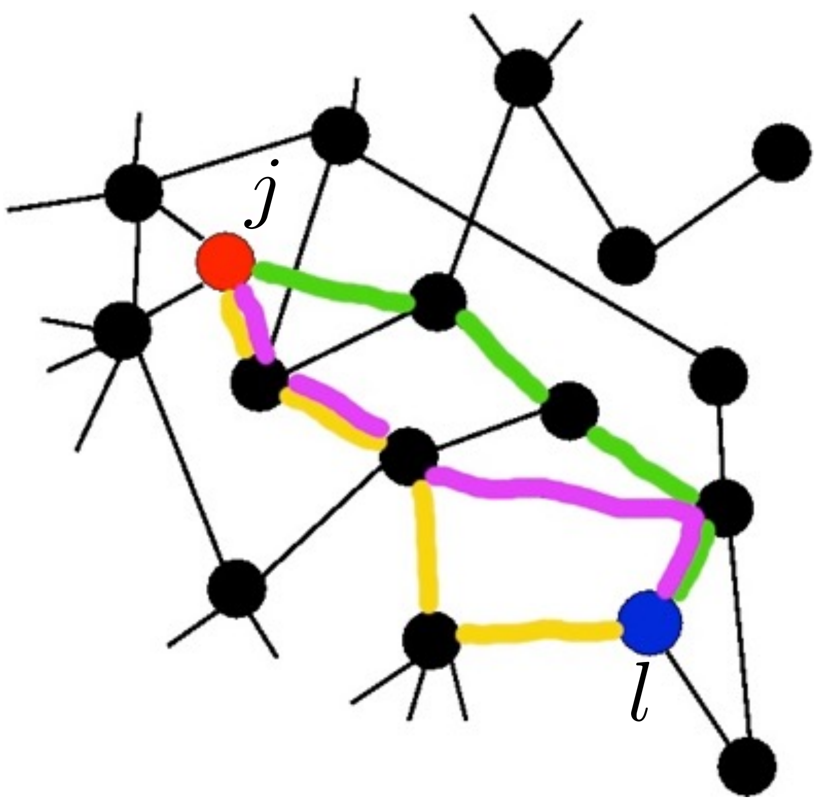


M.A. Beauchamp, Systems Research and Behavioral Science 10, 161 (1965)

II Closeness

Considers the distances between a node and the rest of the network

$$c_j = \frac{1}{N^{-1}\sum_{l=1}^{N} d_{jl}}$$



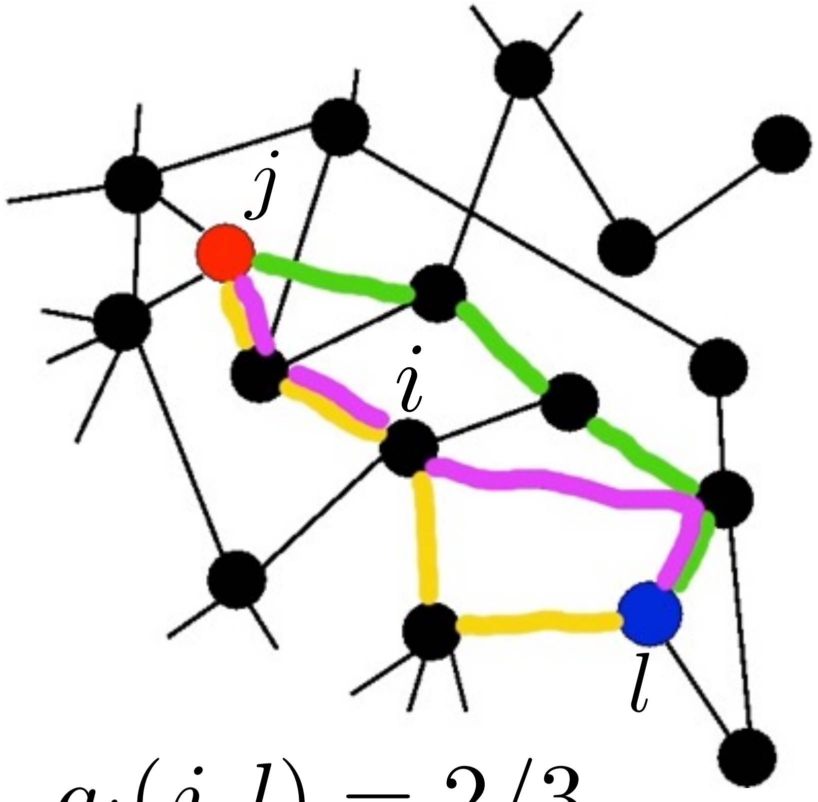
M.A. Beauchamp, Systems Research and Behavioral Science 10, 161 (1965)

@gomezgardenes

III Betweenness

Centrality measure of a node that counts the number of shortest paths that traverse it

$$B_i = \frac{1}{2}\sum_{j,l=1}^{N} g_i(j,l)$$



$$g_i(j,l) = 2/3$$

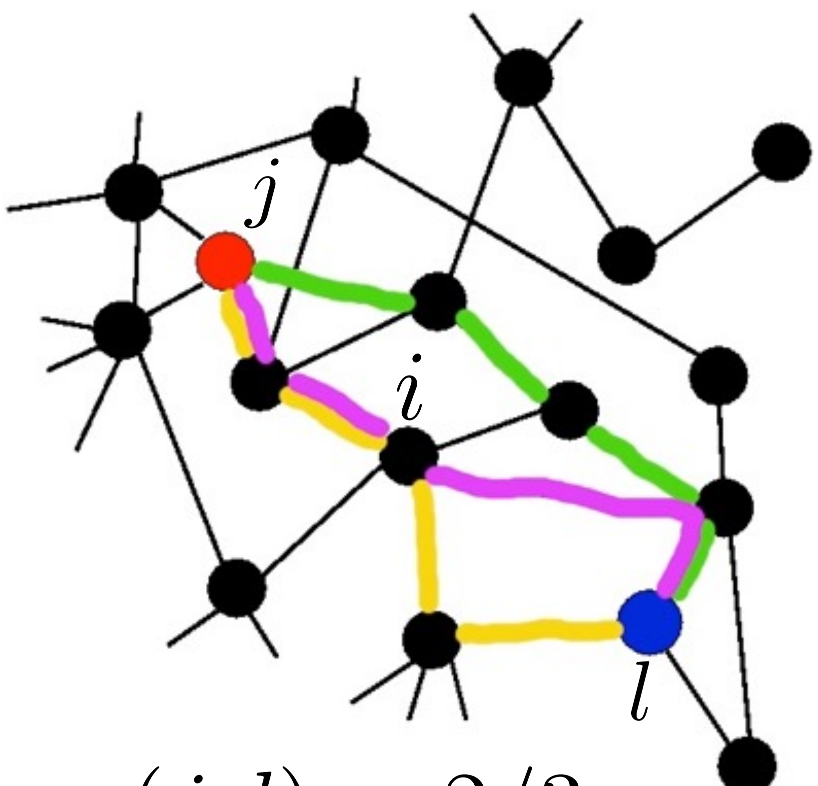
L.C. FREEMAN, SOCIAL NETWORKS 1, 215 (1979)

@gomezgardenes

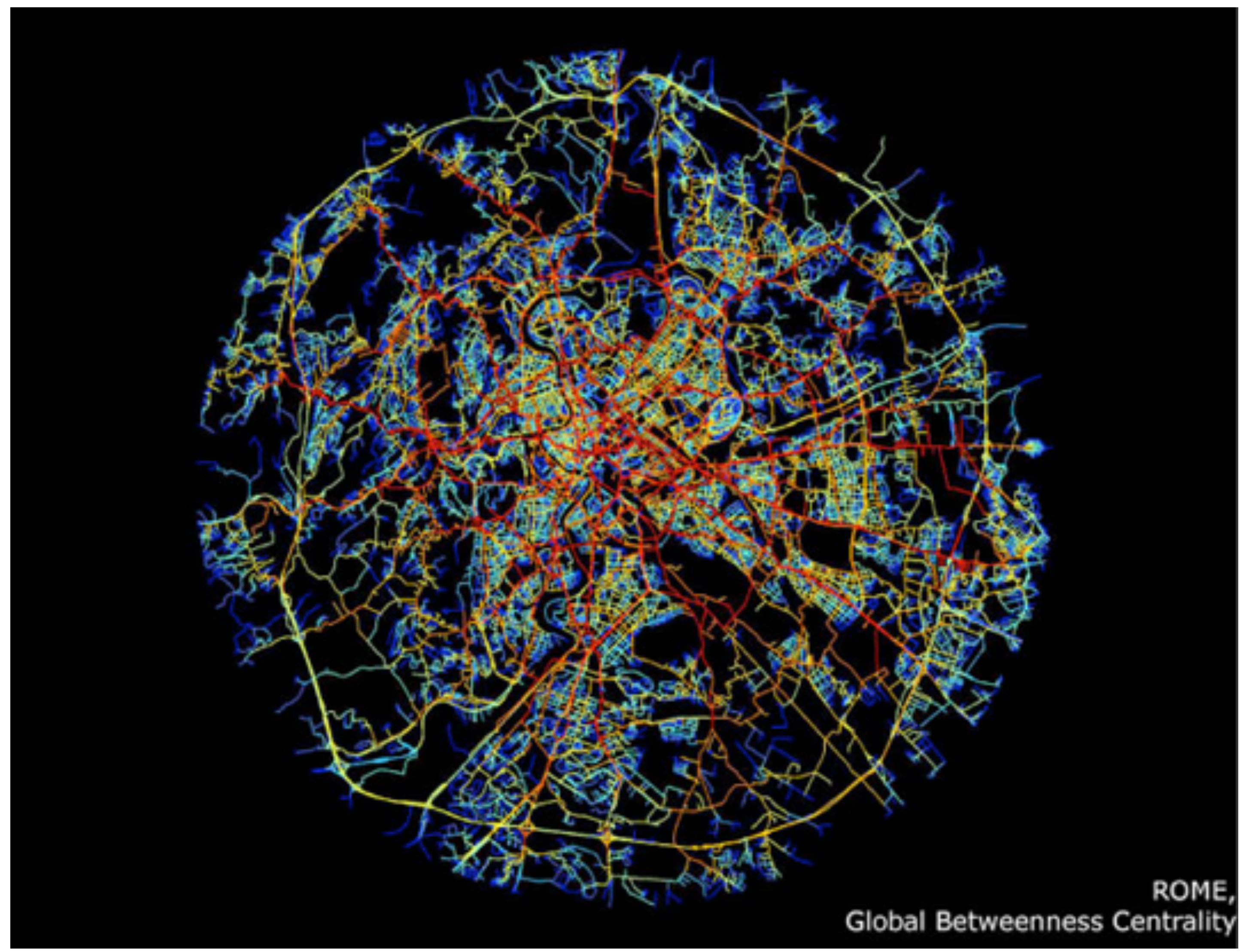
III Betweenness

Centrality measure of a node that counts the number of shortest paths that traverse it

$$B_i = \frac{1}{2}\sum_{j,l=1}^{N} g_i(j,l)$$



$$g_i(j,l) = 2/3$$



L.C. FREEMAN, SOCIAL NETWORKS 1, 215 (1979)

@gomezgardenes

Lets test!

OPEN ACCESS Freely available online

PLOS | ONE

Attack Robustness and Centrality of Complex Networks

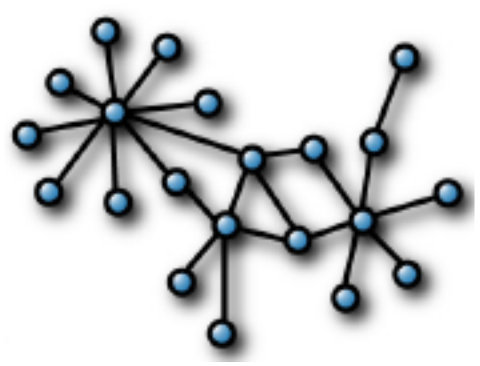
Swami Iyer[1], Timothy Killingback[2]*, Bala Sundaram[3], Zhen Wang[4]

1 Computer Science Department, University of Massachusetts, Boston, Massachusetts, United States of America, **2** Mathematics Department, University of Massachusetts, Boston, Massachusetts, United States of America, **3** Physics Department, University of Massachusetts, Boston, Massachusetts, United States of America, **4** Physics Department, University of Massachusetts, Boston, Massachusetts, United States of America

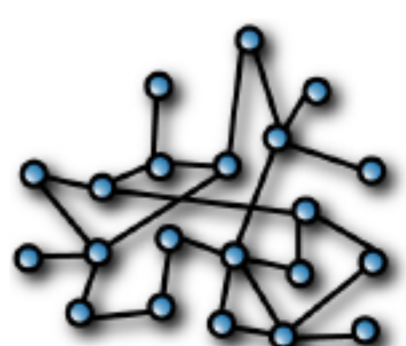
Simultaneous

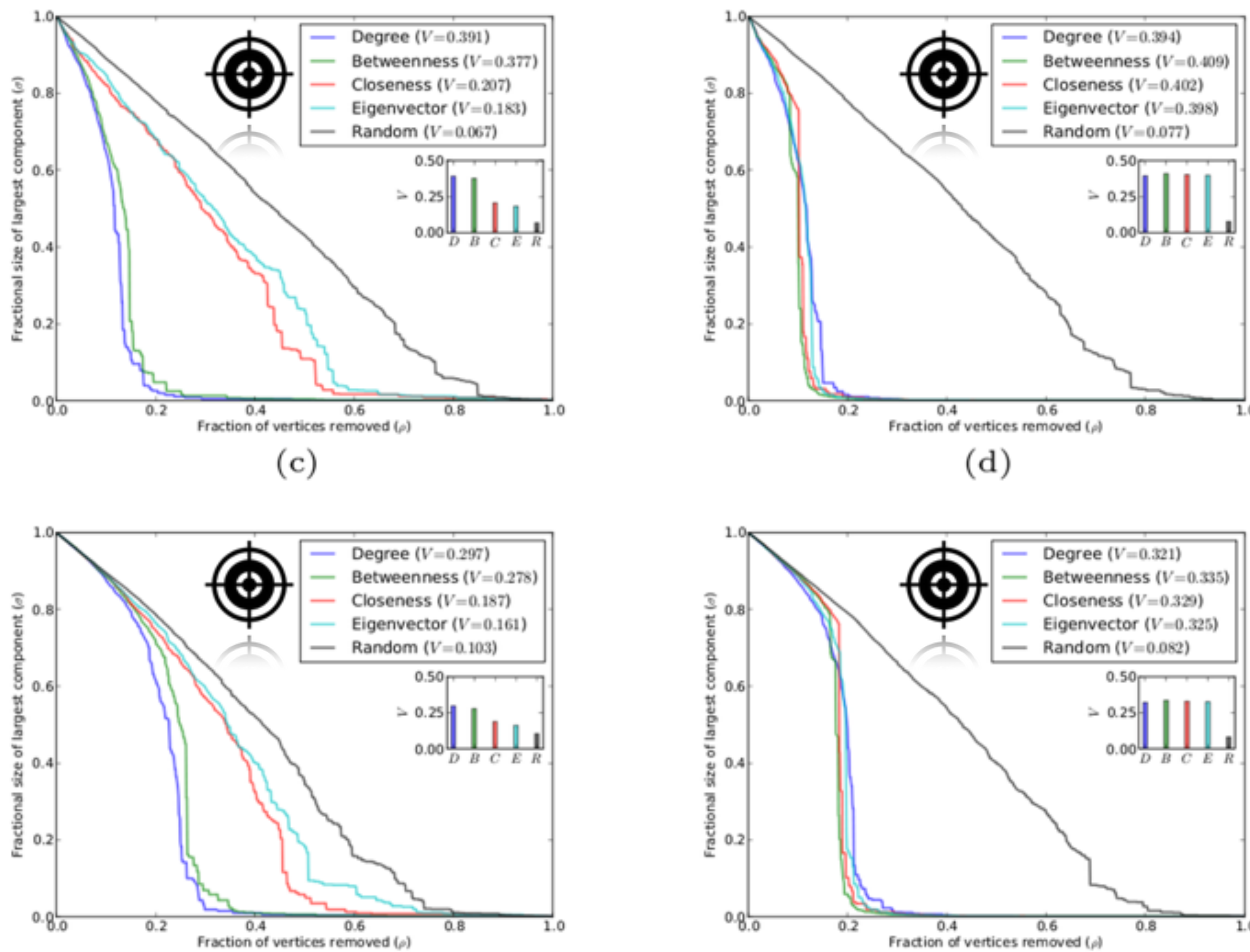
Sequential

Scale-free network



Random Graph





OPEN ACCESS Freely available online

PLOS | ONE

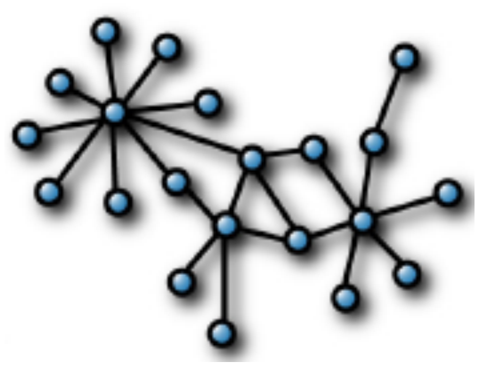
Attack Robustness and Centrality of Complex Networks

Swami Iyer[1], Timothy Killingback[2]*, Bala Sundaram[3], Zhen Wang[4]

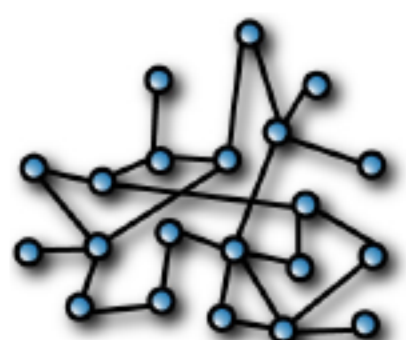
1 Computer Science Department, University of Massachusetts, Boston, Massachusetts, United States of America, 2 Mathematics Department, University of Massachusetts, Boston, Massachusetts, United States of America, 3 Physics Department, University of Massachusetts, Boston, Massachusetts, United States of America, 4 Physics Department, University of Massachusetts, Boston, Massachusetts, United States of America

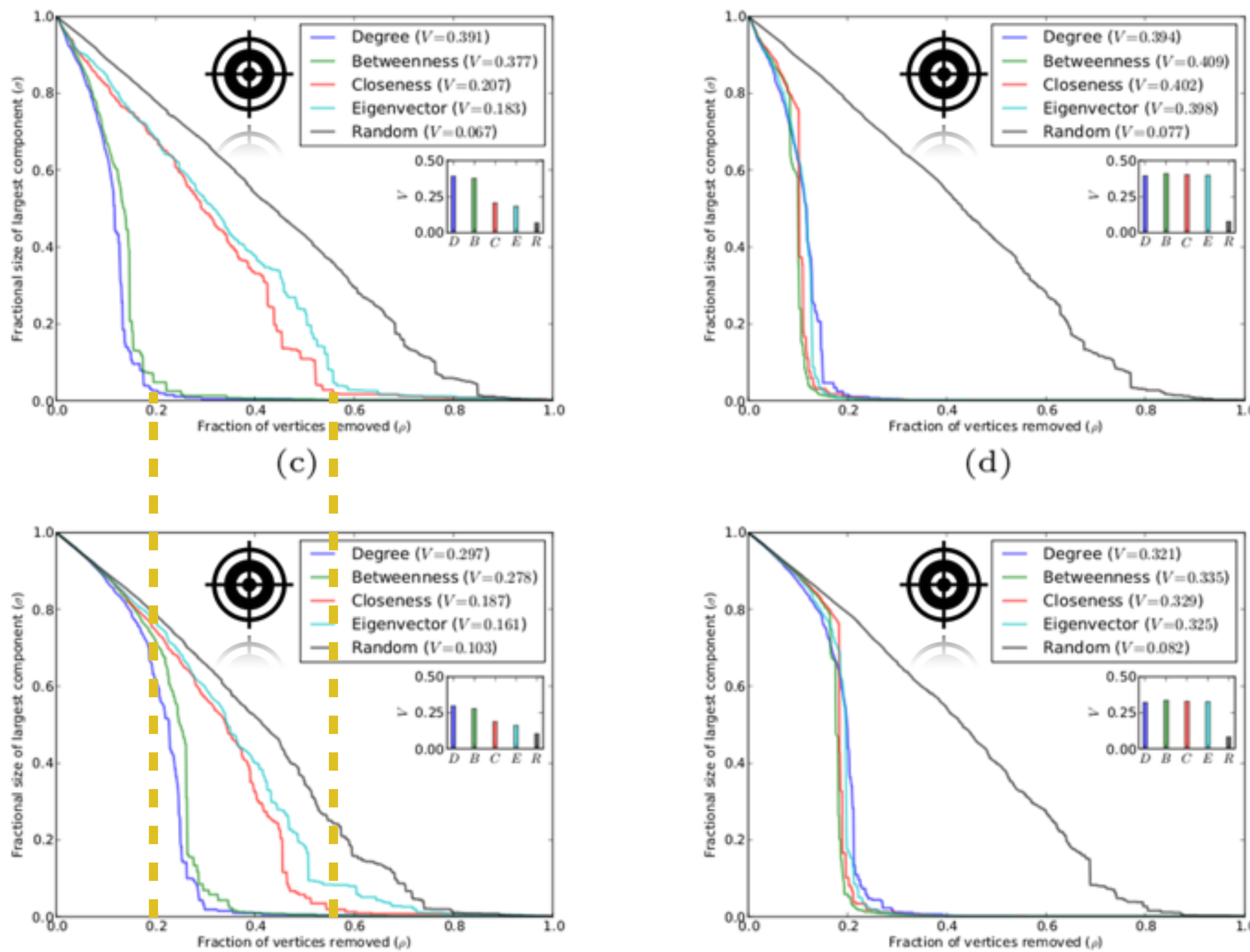
Simultaneous | Sequential

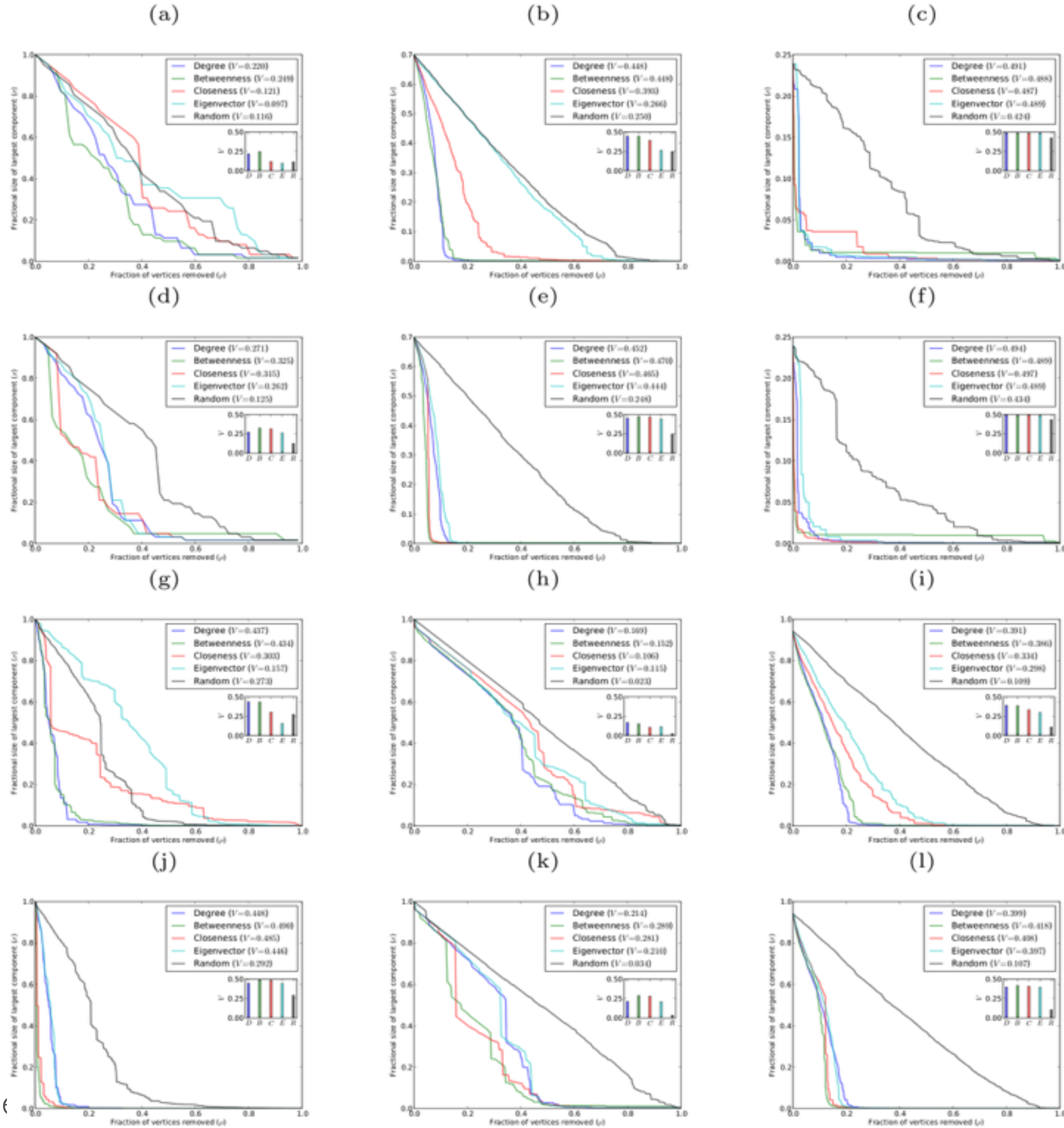
Scale-free network

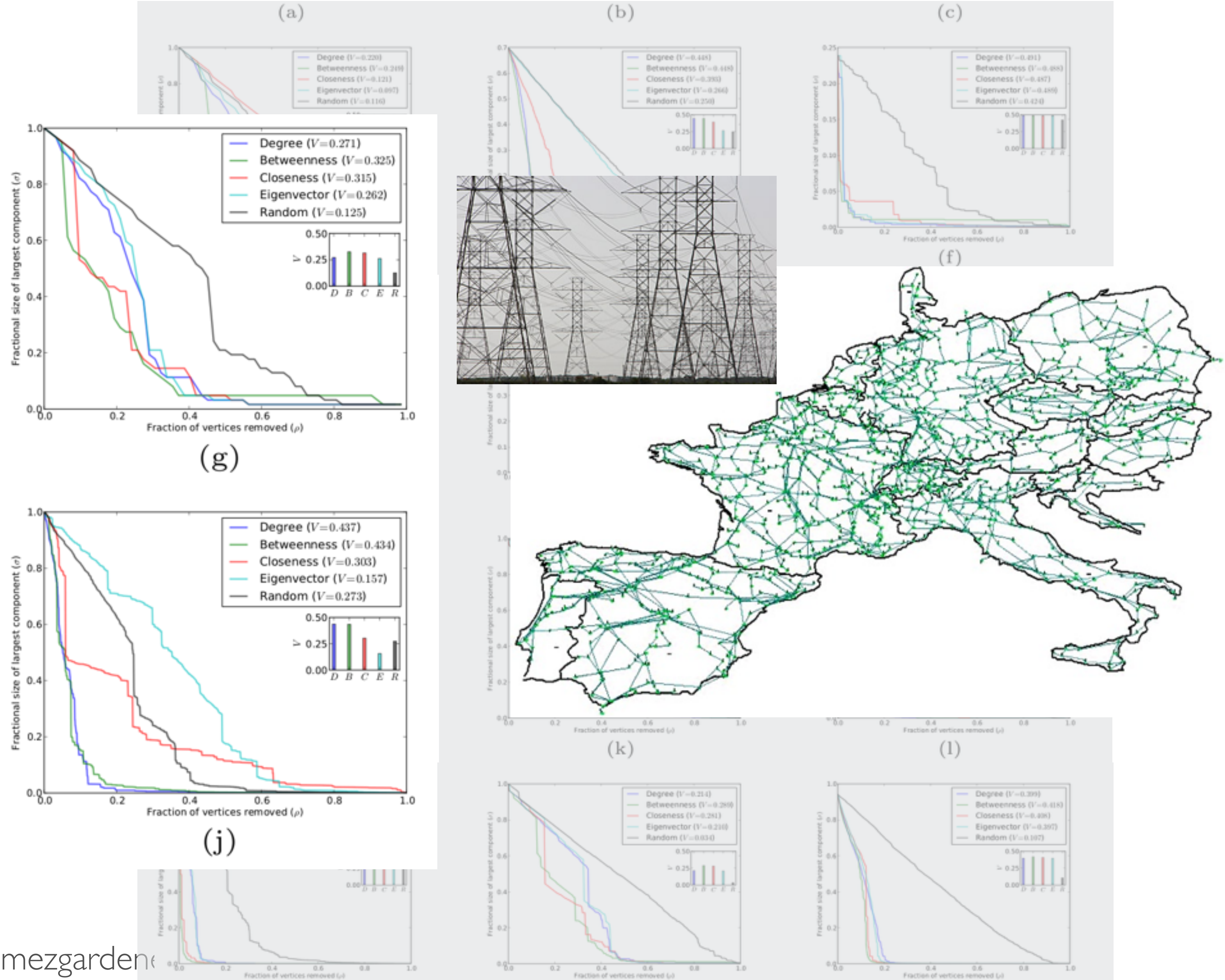


Random Graph

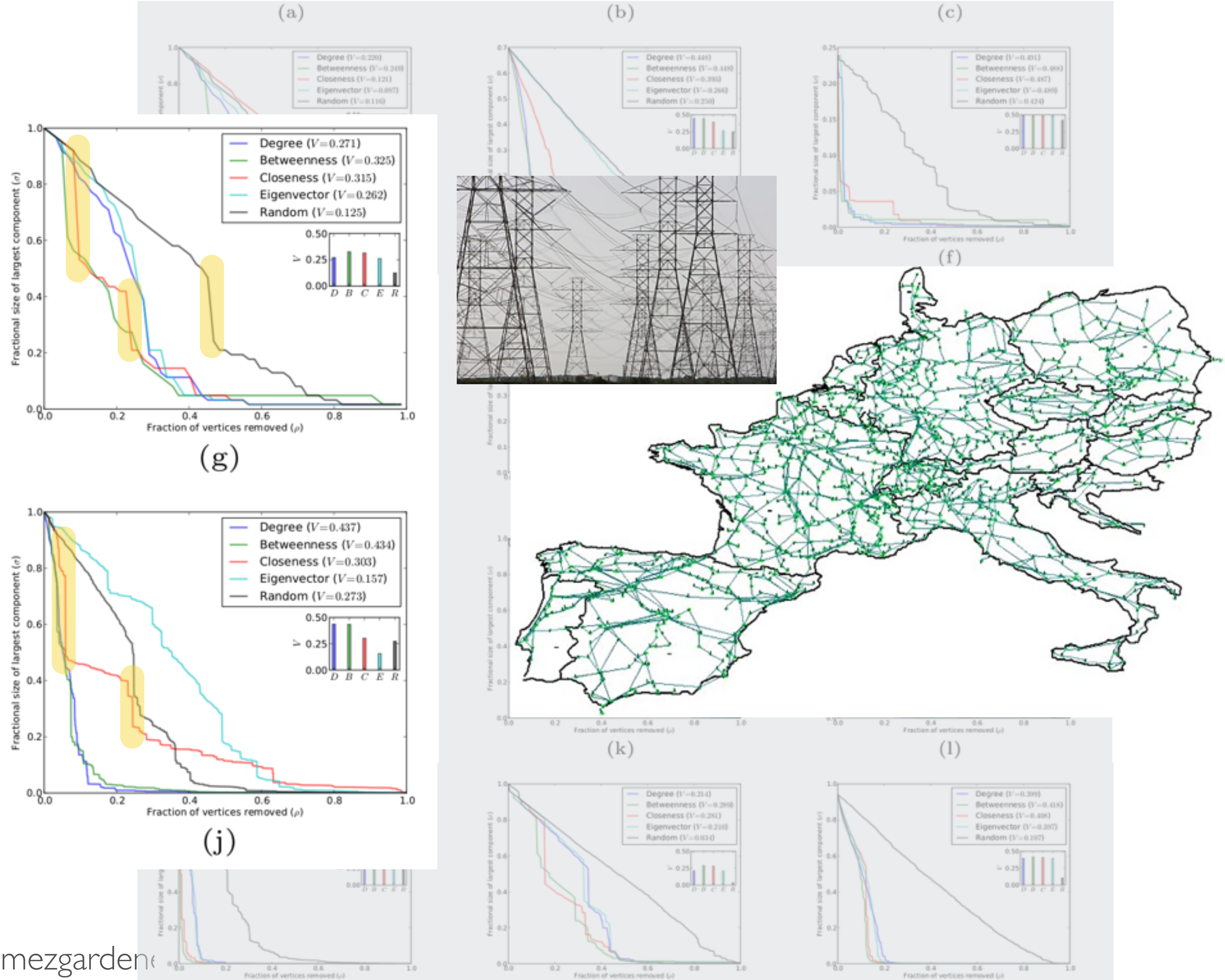




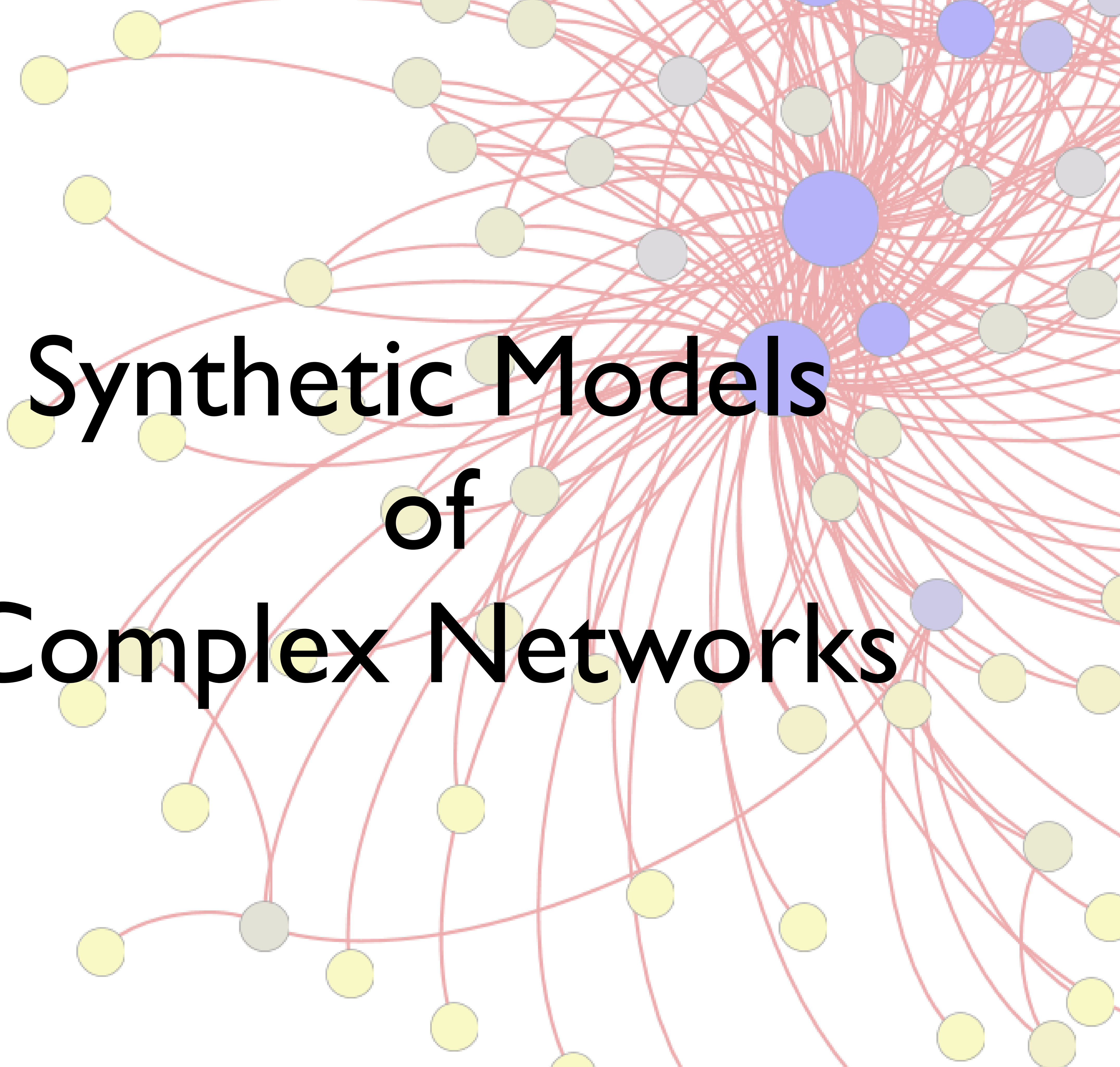

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
Degree
Betweenness
Closeness
Eigenvector
Random
Fraction of vertices removed
Fractional size of largest component



@gomezgarden



@gomezgardene



Synthetic Models of Complex Networks

@gomezgardenes

Two frameworks

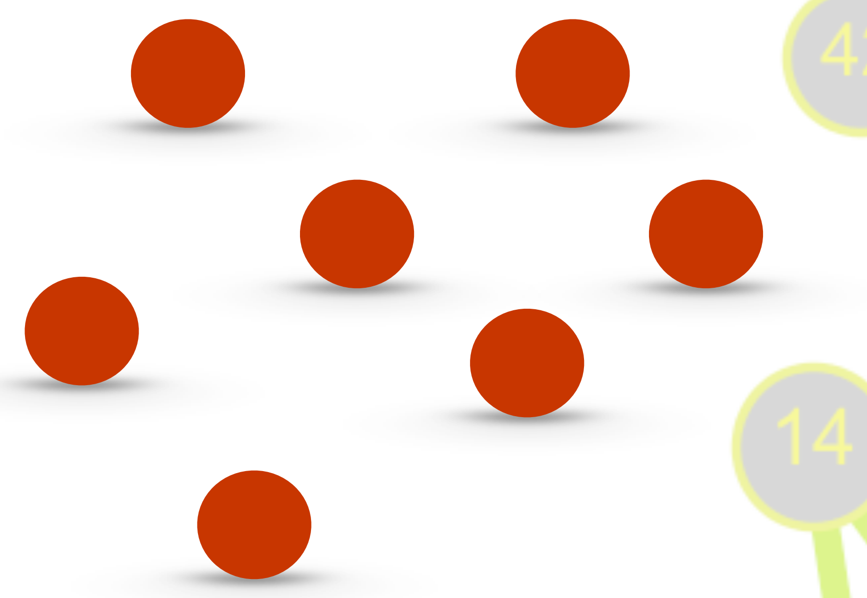
- Equilibrium Random Networks
- Non-equilibrium Random Networks

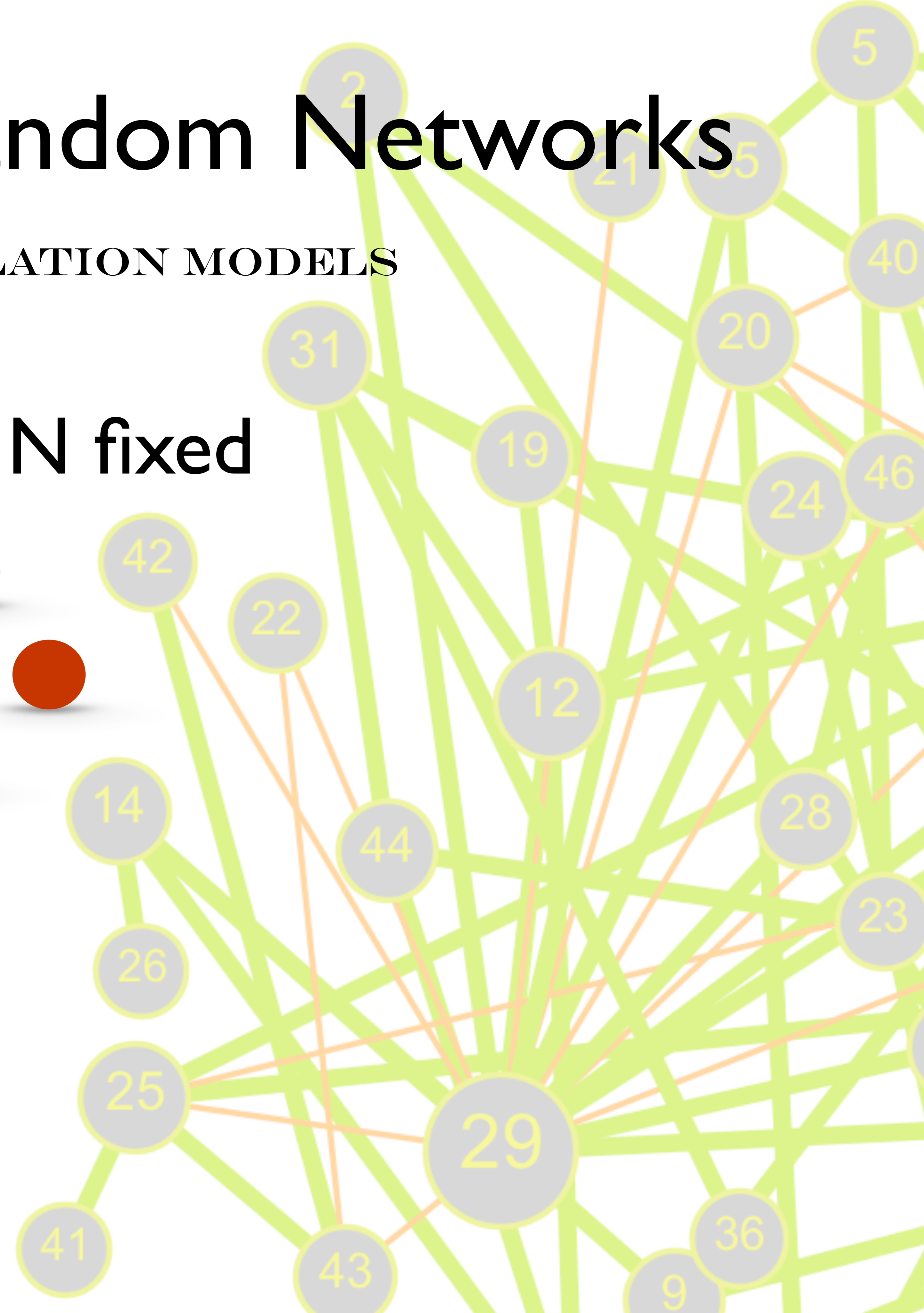
@gomezgardenes

Equilibrium Random Networks

PERCOLATION MODELS

- Number of nodes N fixed



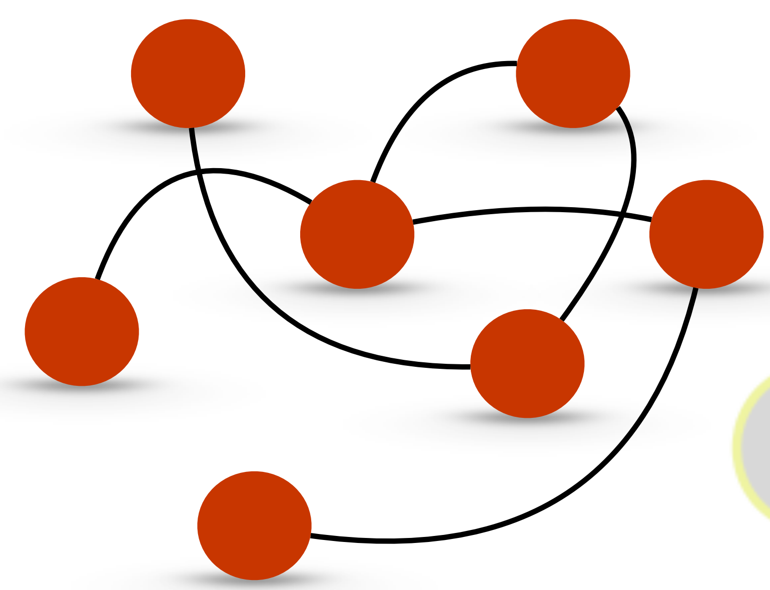


@gomezgardenes

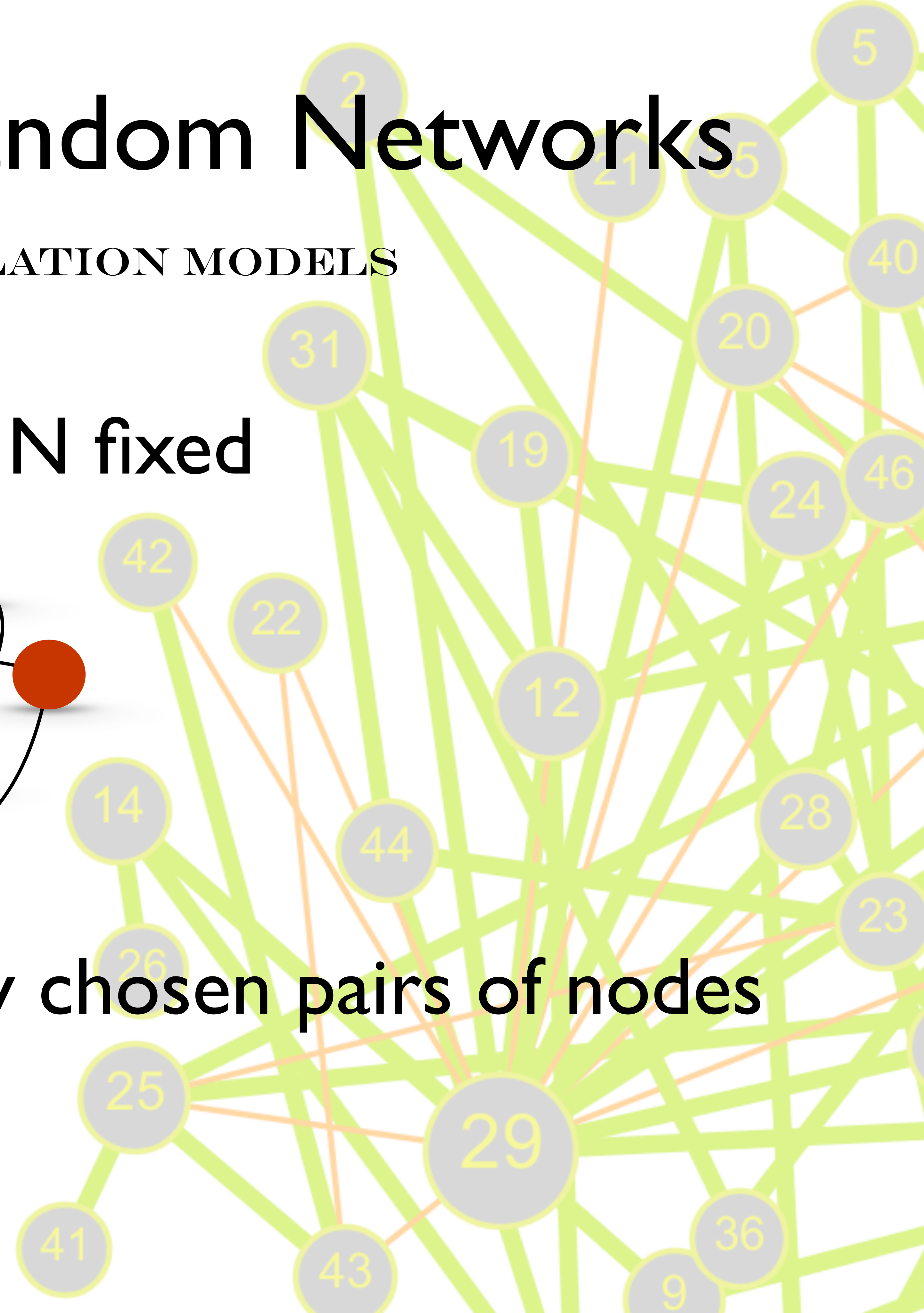
Equilibrium Random Networks

PERCOLATION MODELS

- Number of nodes N fixed



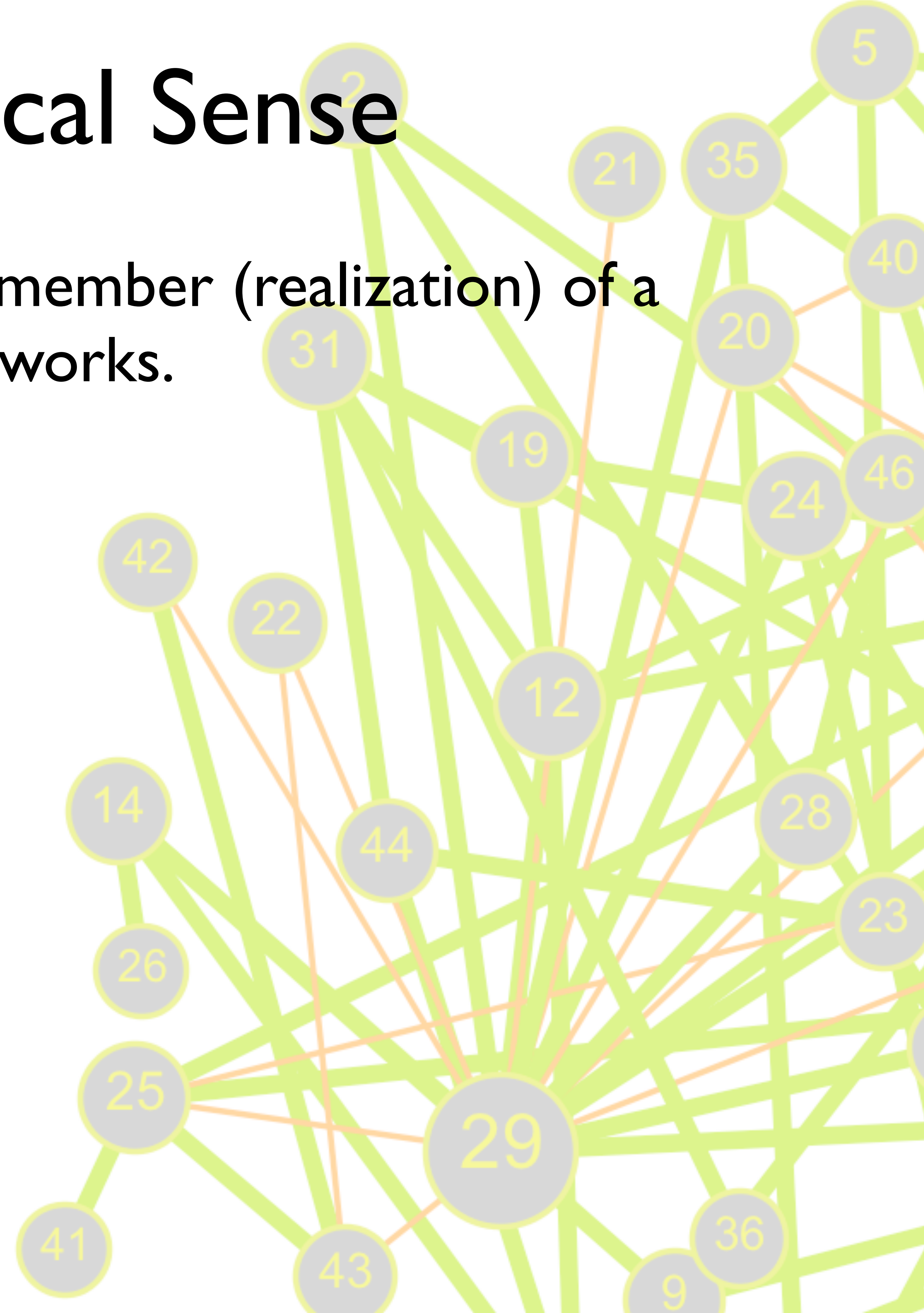
- Connect randomly chosen pairs of nodes



@gomezgardenes

Statistical Sense

- A particular network is a member (realization) of a statistical ensemble of networks.

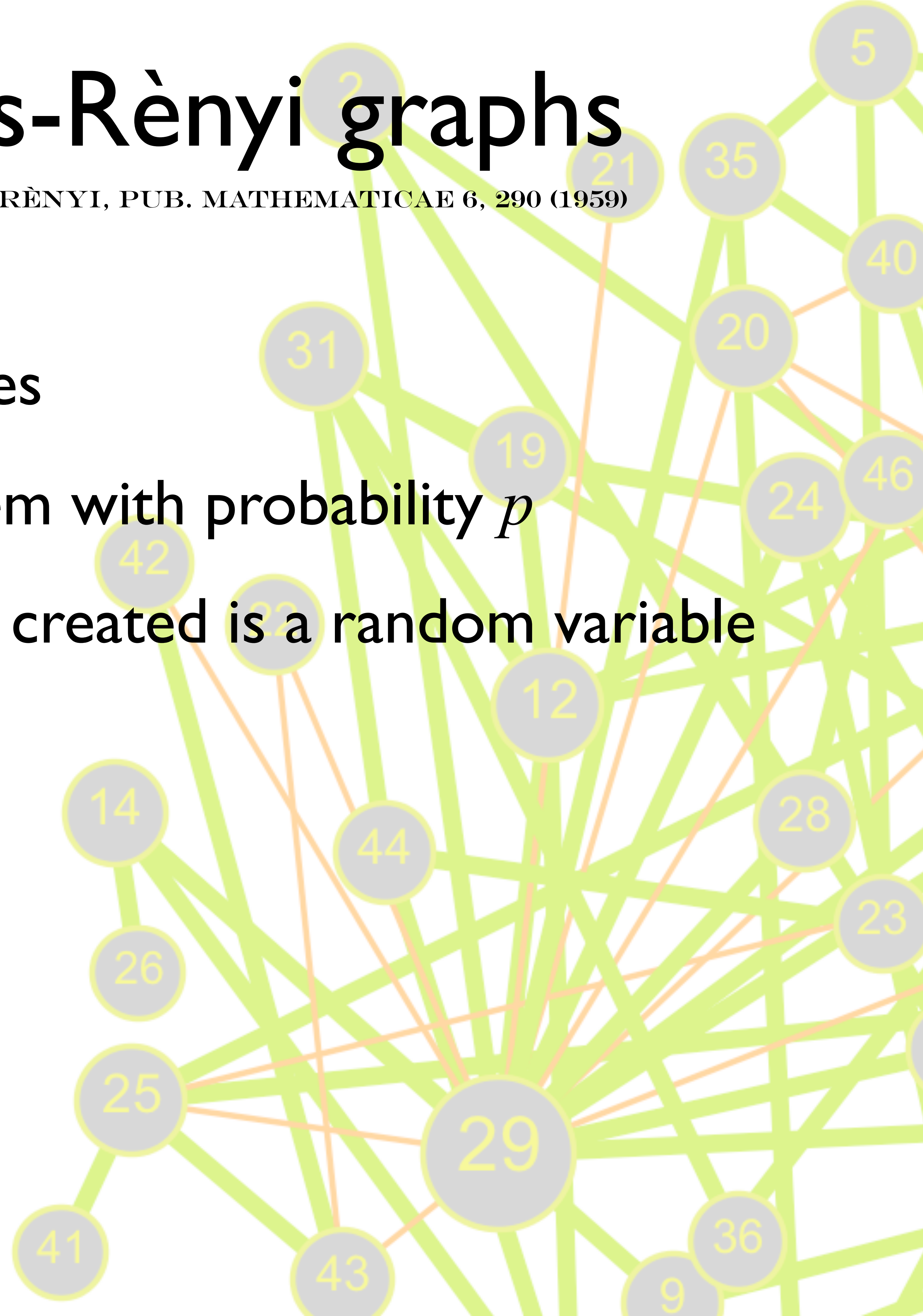


@gomezgardenes

Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

- Start with N isolated nodes
- For each pair connect them with probability p
- The total number of links created is a random variable



Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

- Start with N isolated nodes
- For each pair connect them with probability p
- The total number of links created is a random variable

$$\boxed{E(L) = p\frac{N(N-1)}{2}}$$

@gomezgardenes

Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

- Start with N isolated nodes
- For each pair connect them with probability p
- The total number of links created is a random variable

$$\boxed{E(L) = p\frac{N(N-1)}{2}}$$

- The probability of finding graphs with L links is

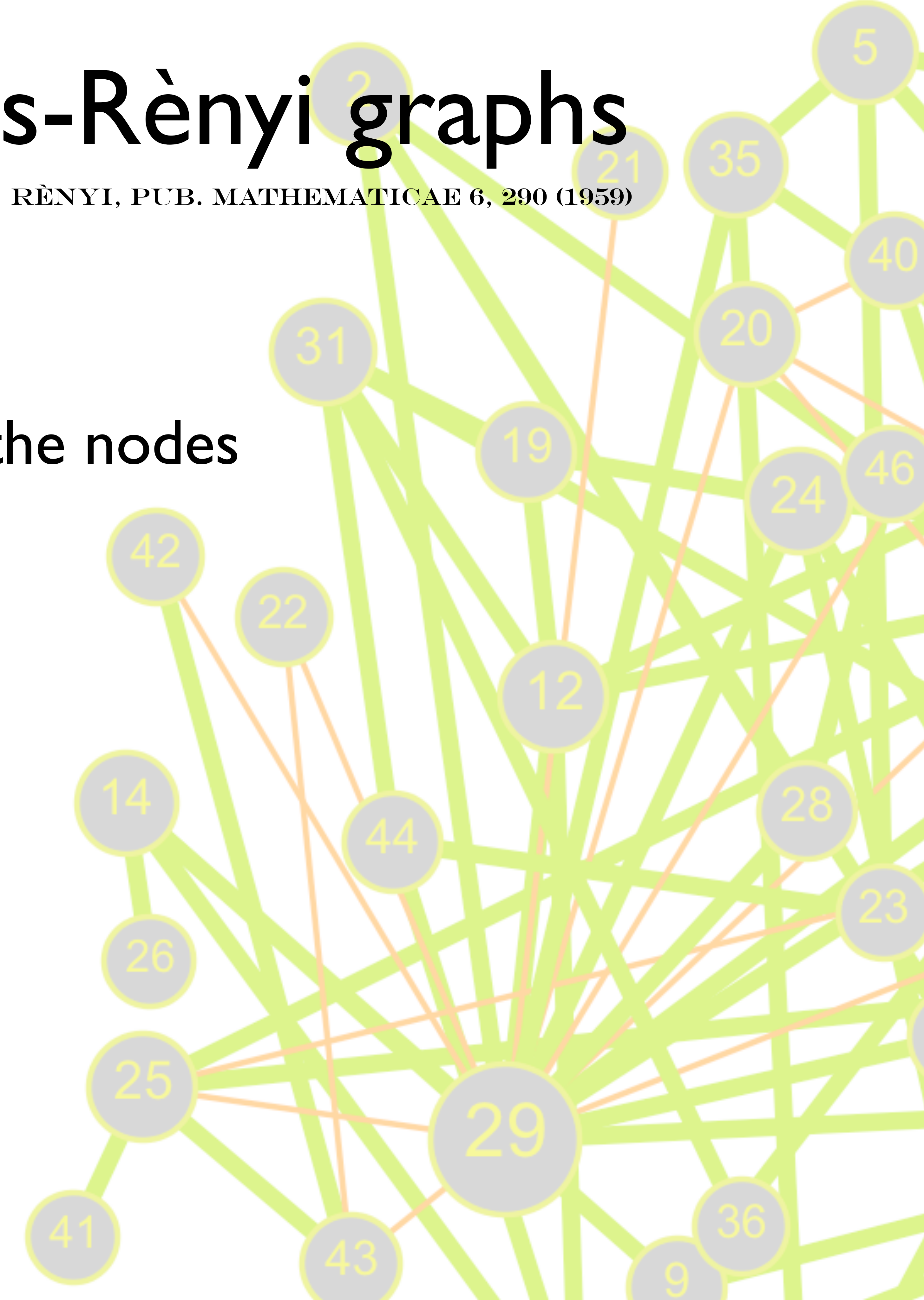
$$\boxed{P(N,L) = \binom{N(N-1)/2}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}}$$

@gomezgardenes

Erdös-Rènyi graphs

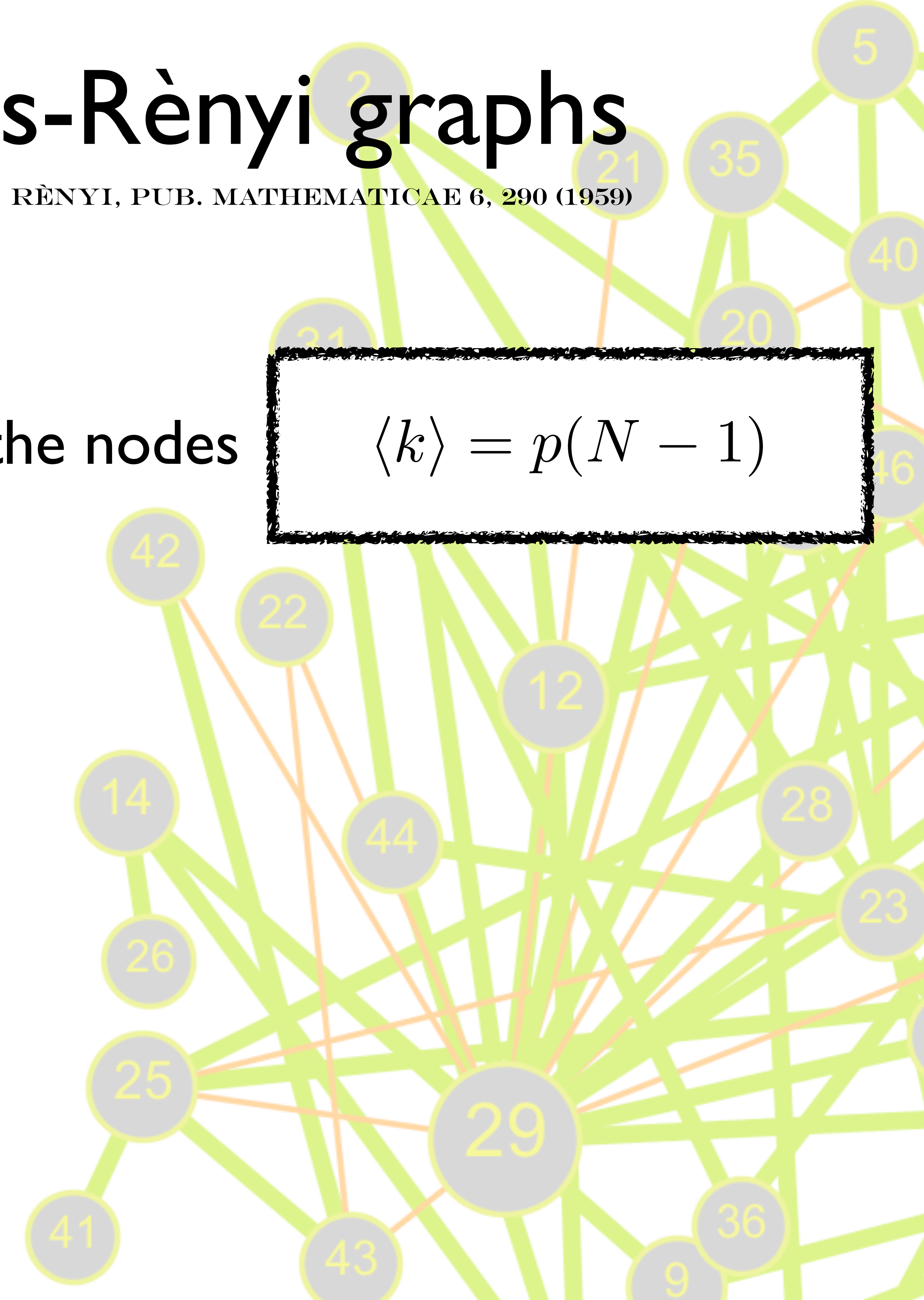
P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

- Average connectivity of the nodes



Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)



- Average connectivity of the nodes $$\boxed{\langle k \rangle = p(N-1)}$$

@gomezgardenes

Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

- Average connectivity of the nodes

$$\boxed{\langle k\rangle = p(N-1)}$$

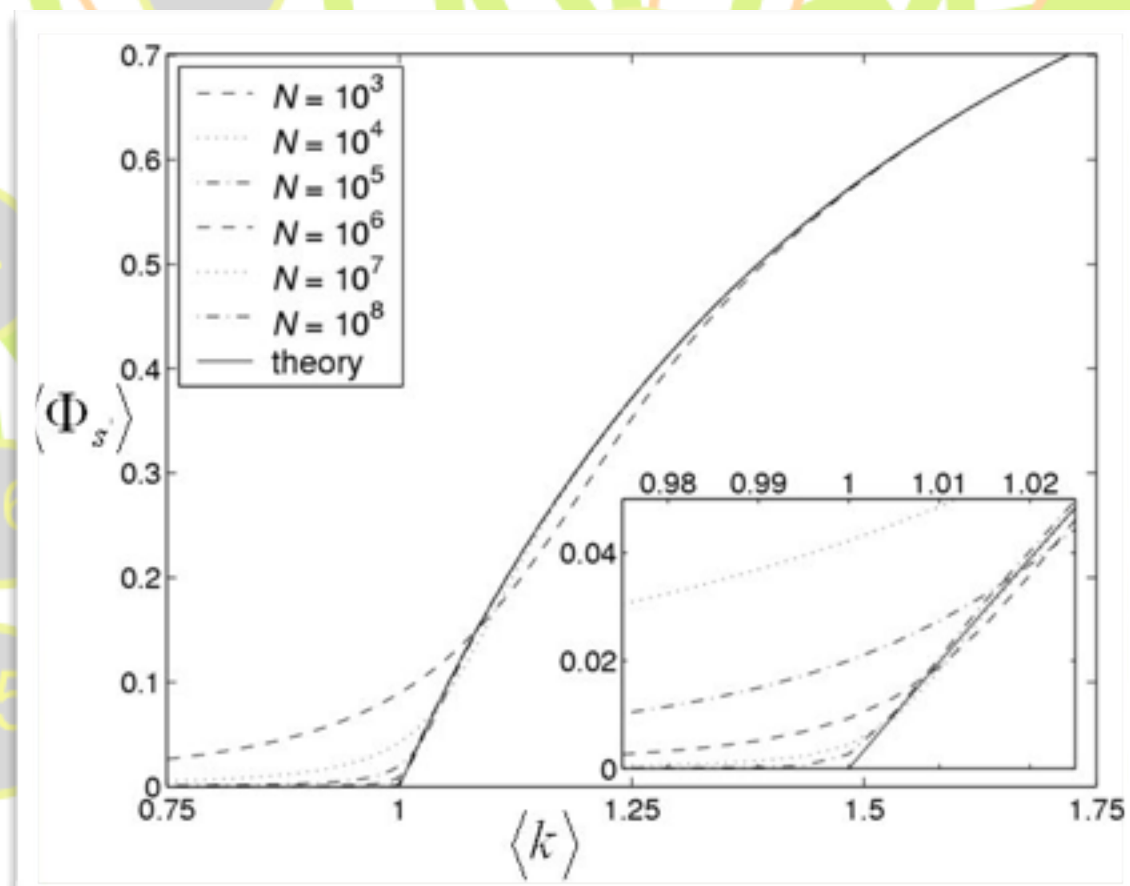
- Percolation Transition:

For $\langle k \rangle < 1$:

Isolated clusters

For $\langle k \rangle > 1$:

Giant Connected Component appears



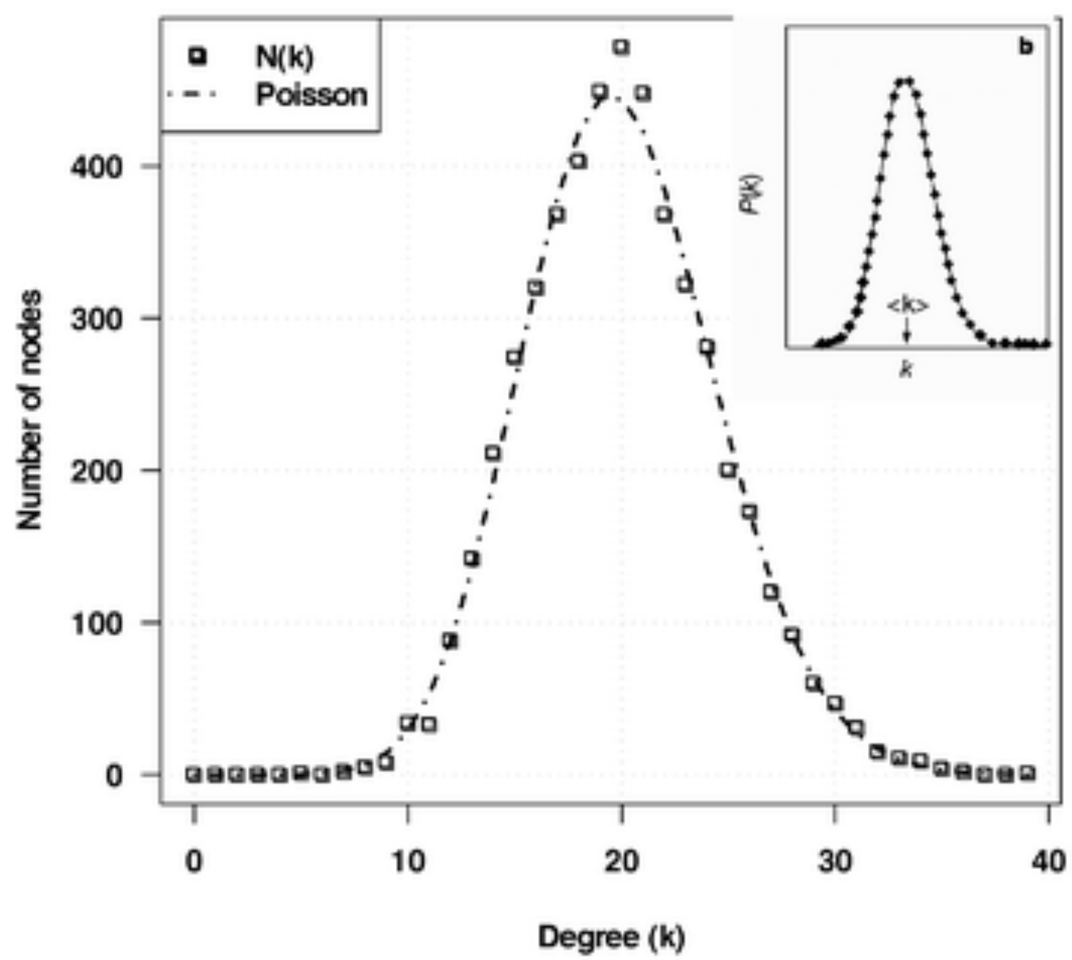
@gomezgardenes

Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

Poisson Degree distribution

$$P(k) \simeq \mathrm{e}^{-pN} \frac{(pN)^k}{k!} = \mathrm{e}^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

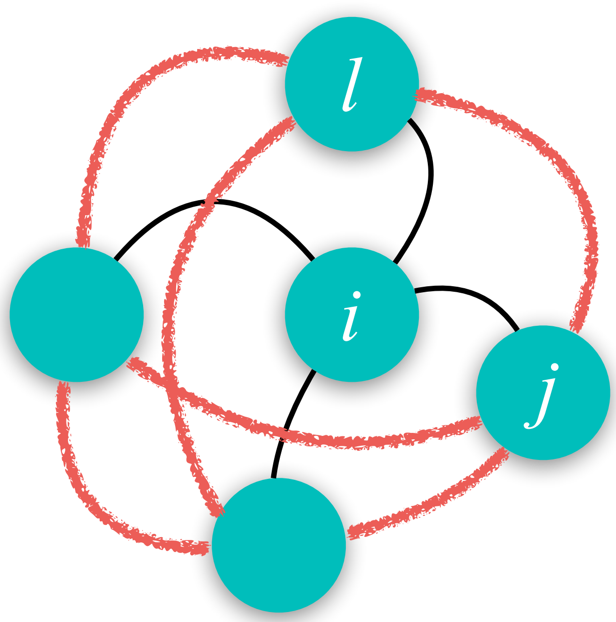


@gomezgardenes

Erdös-Rènyi graphs

P. ERDÖS & A. RÈNYI, PUB. MATHEMATICAE 6, 290 (1959)

Clustering Coefficient



Probability that two nodes j and l are connected, provided they are both connected to a third one i, is: p

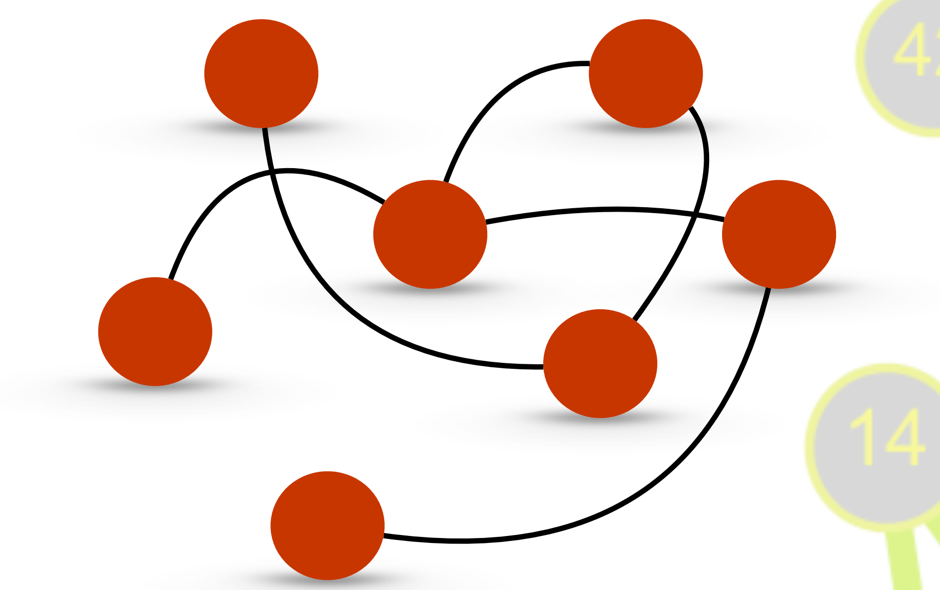
$$C = p = \frac{\langle k \rangle}{N}$$

Clustering tends to 0 as N increases!!!

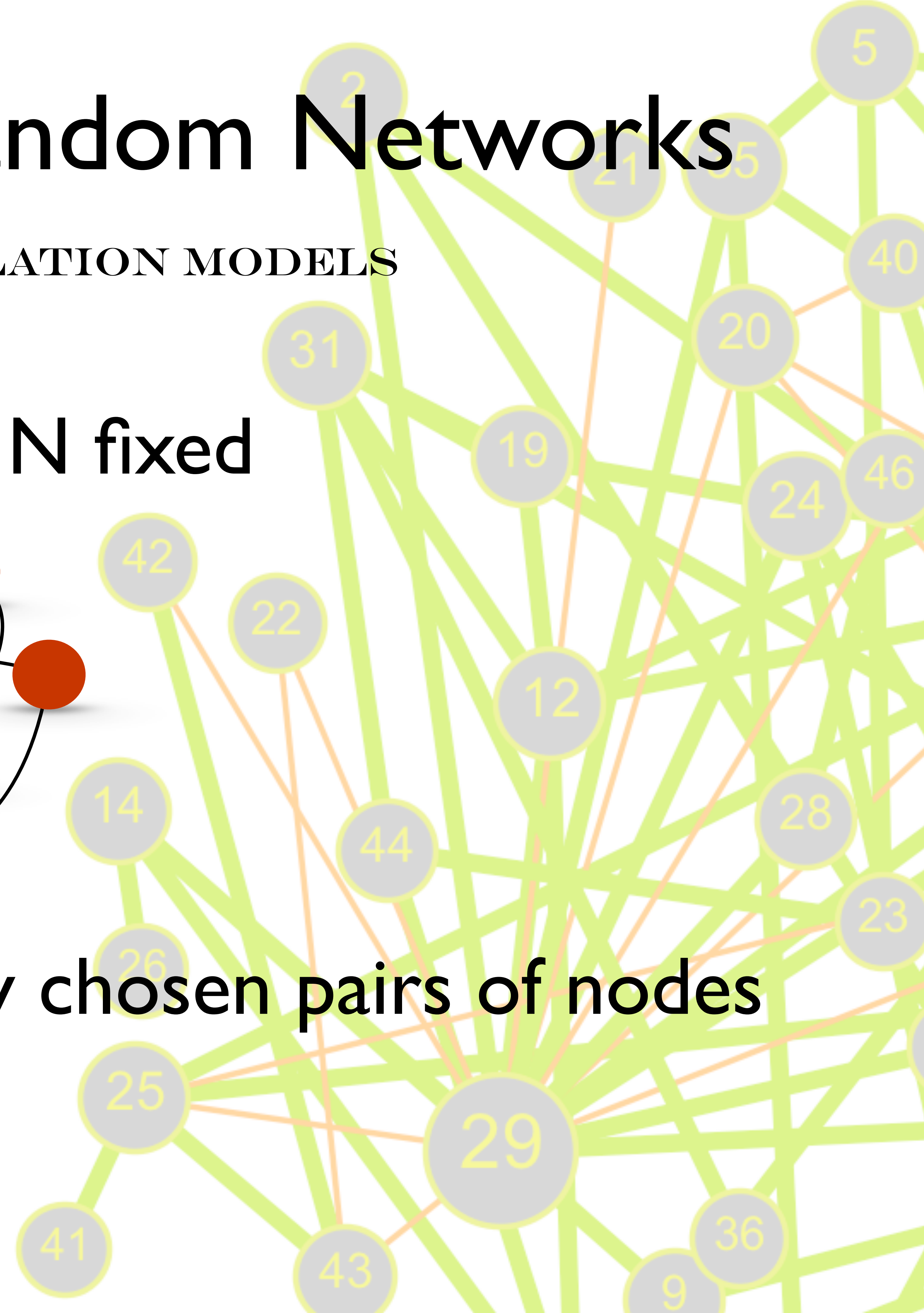
Equilibrium Random Networks

PERCOLATION MODELS

- Number of nodes N fixed



- Connect randomly chosen pairs of nodes

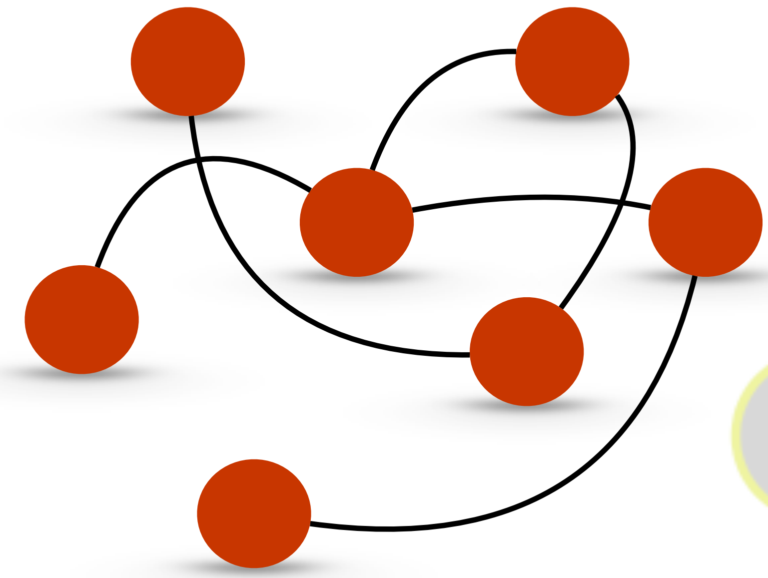


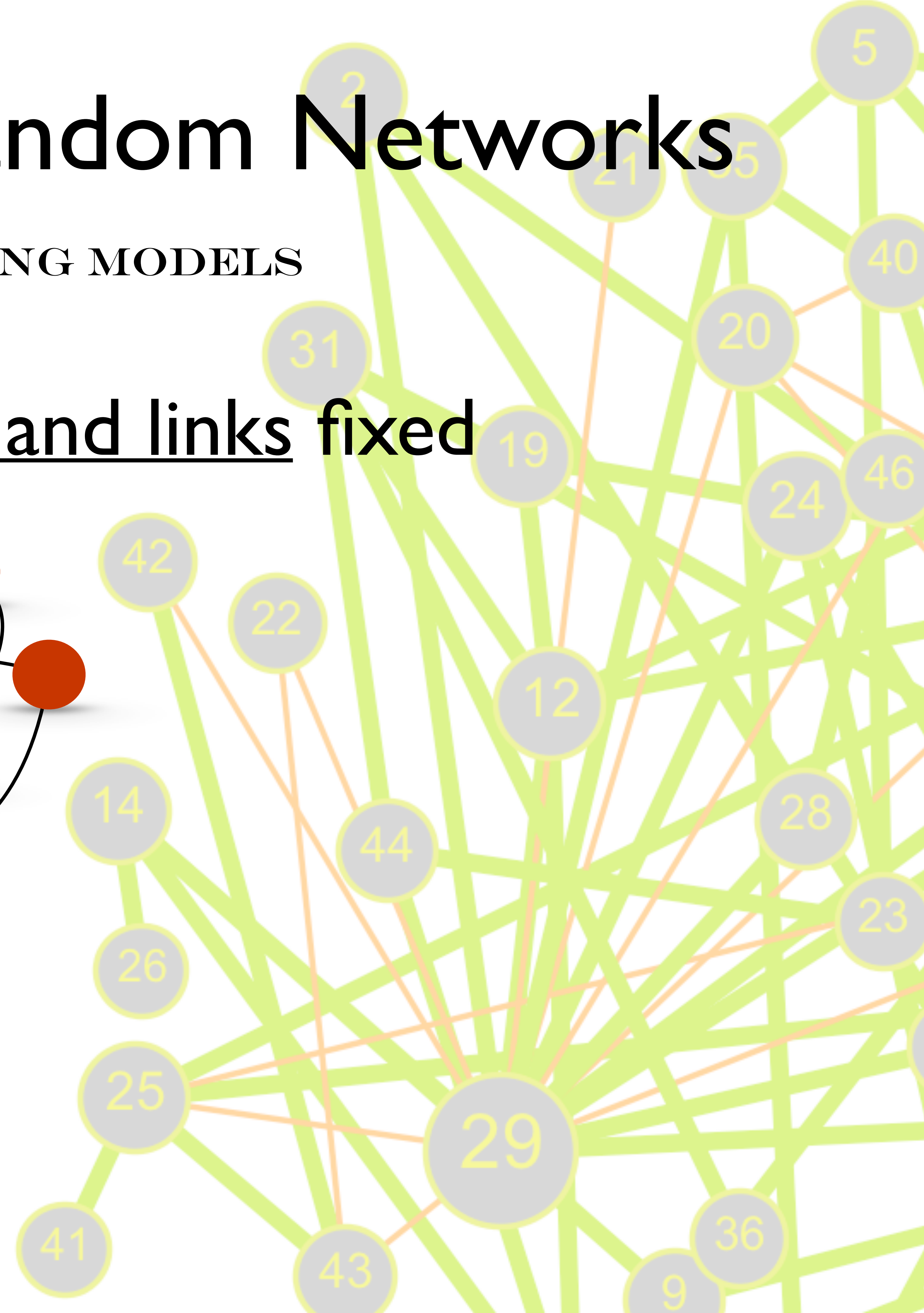
@gomezgardenes

Equilibrium Random Networks

REWIRING MODELS

- Number of <u>nodes and links</u> fixed



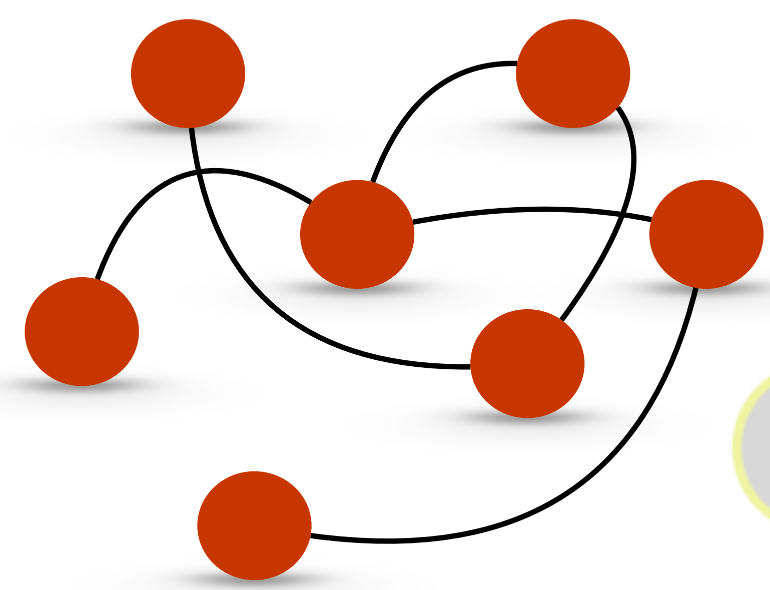


@gomezgardenes

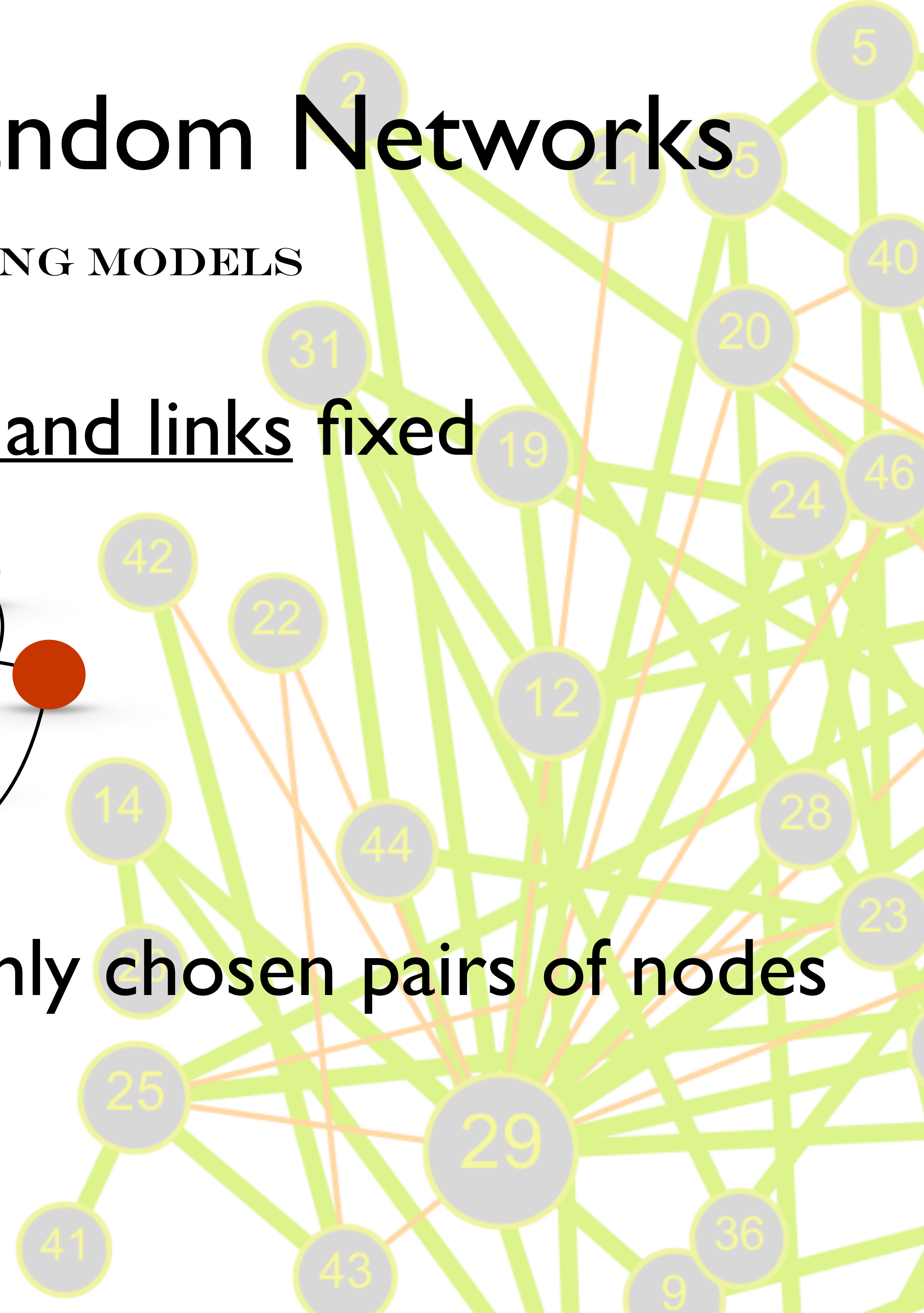
Equilibrium Random Networks

REWIRING MODELS

- Number of nodes and links fixed



- Reconnect randomly chosen pairs of nodes

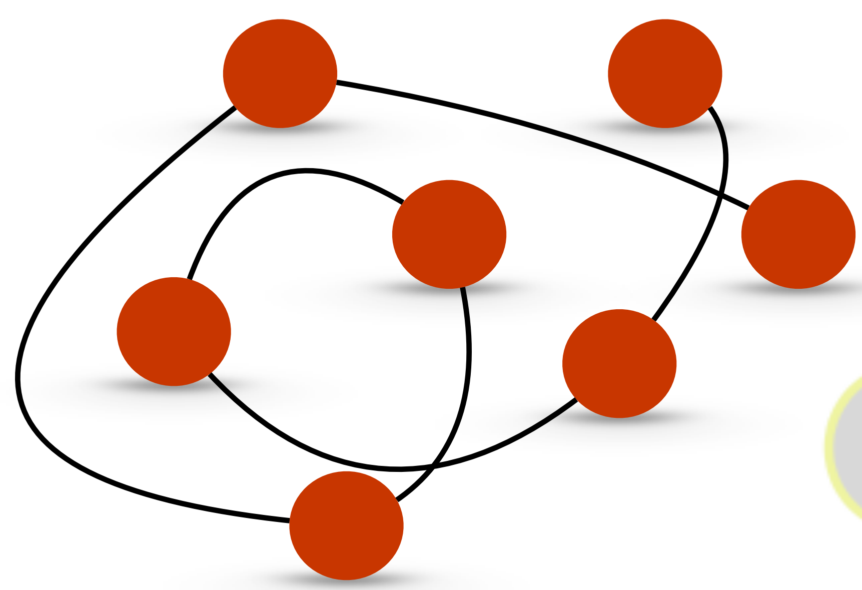


@gomezgardenes

Equilibrium Random Networks

REWIRING MODELS

- Number of nodes and links fixed



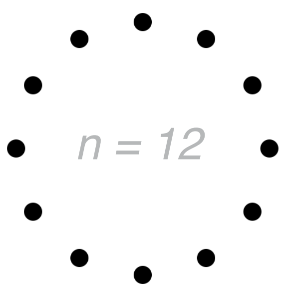
- Reconnect randomly chosen pairs of nodes

@gomezgardenes

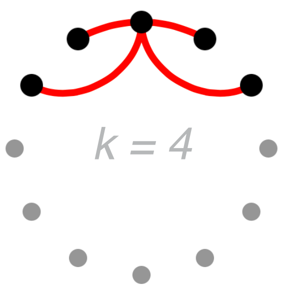
Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

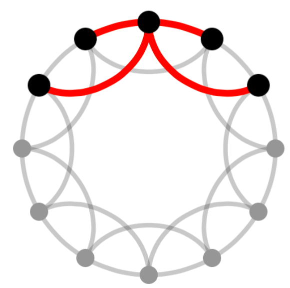
We start with a ring of *n* vertices



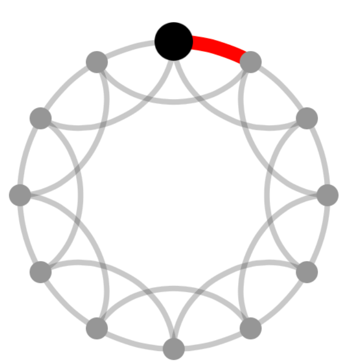
where each vertex is connected to its *k* nearest neighbors



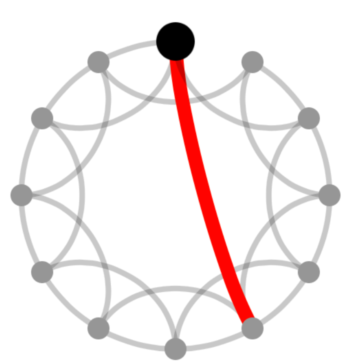
like so.



We choose a vertex, and the edge to its nearest clockwise neighbour.



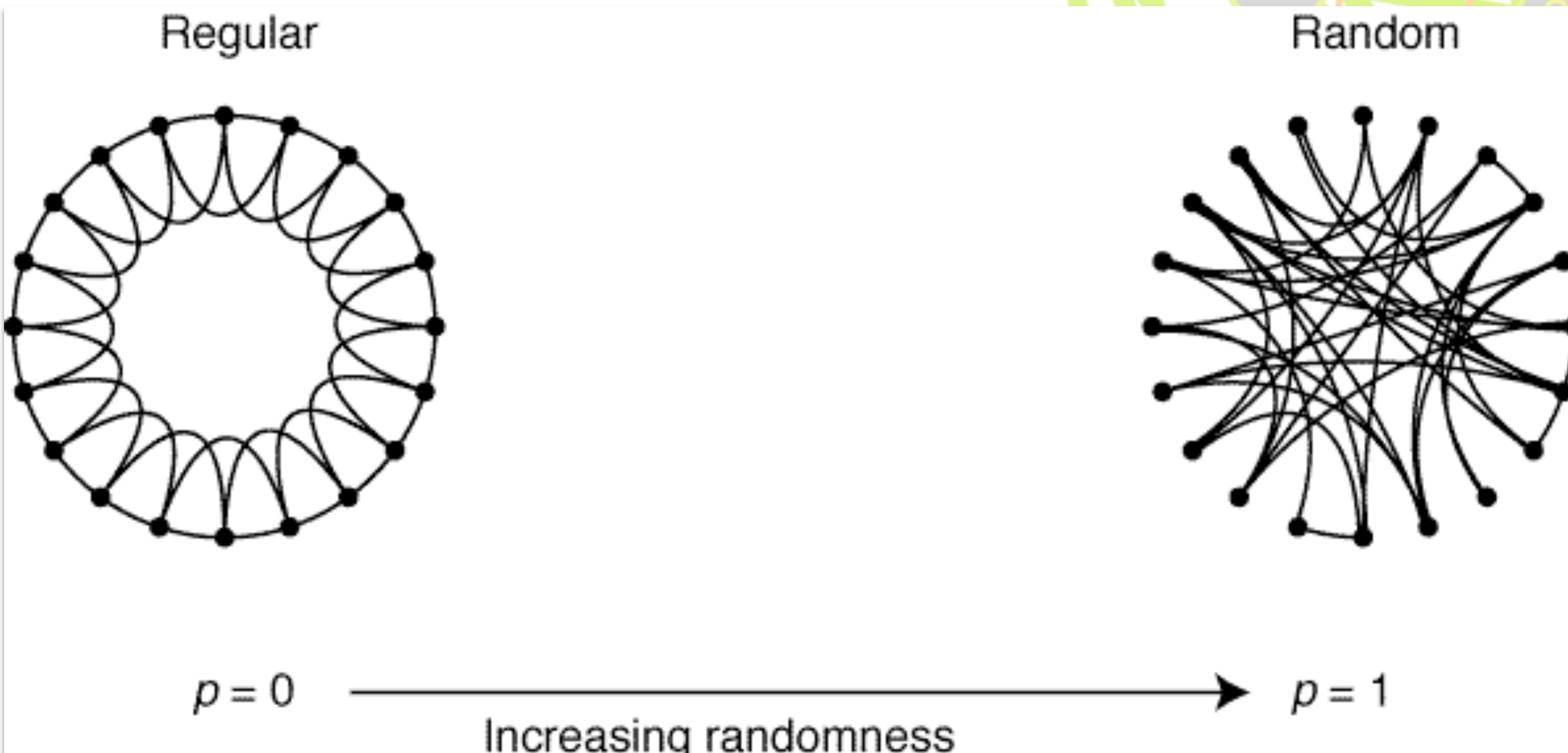
With probability *p*, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden. Otherwise, we leave the edge in place.



@gomezgardenes

Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

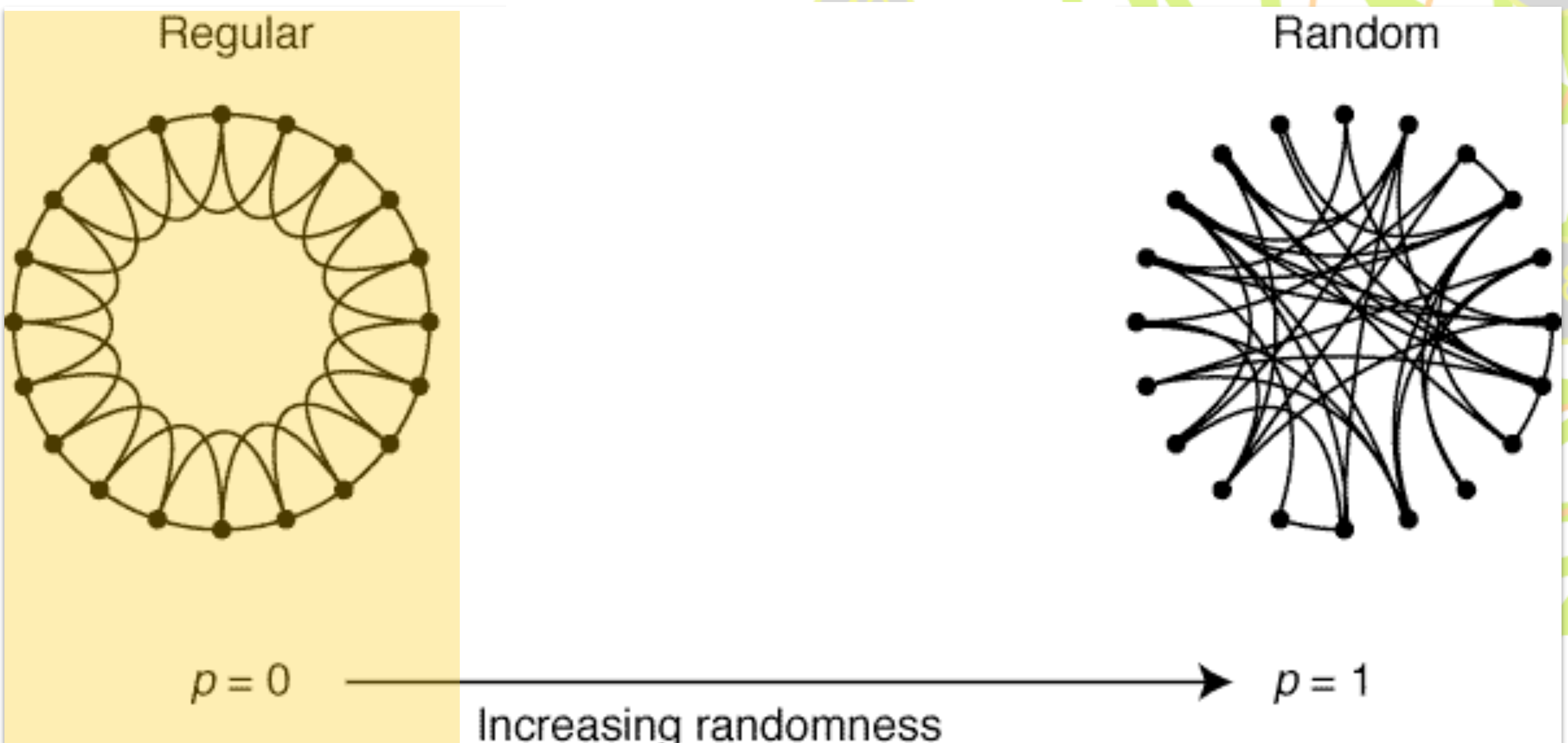


@gomezgardenes

Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

$$L \simeq \frac{N}{2\langle k\rangle} \gg 1 \qquad\qquad C = \frac{3(\langle k\rangle - 2)}{4(\langle k\rangle - 1)}$$

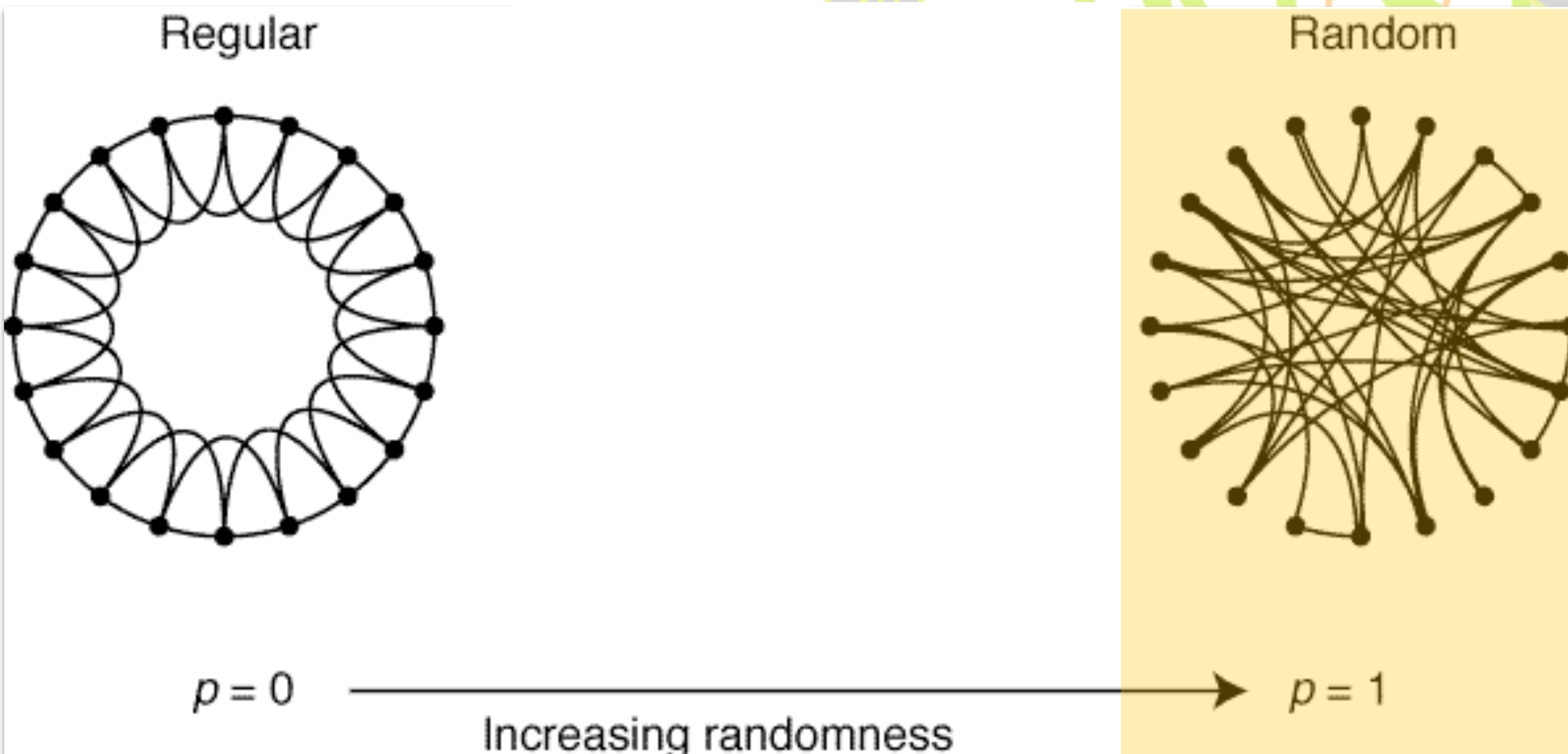


@gomezgardenes

Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

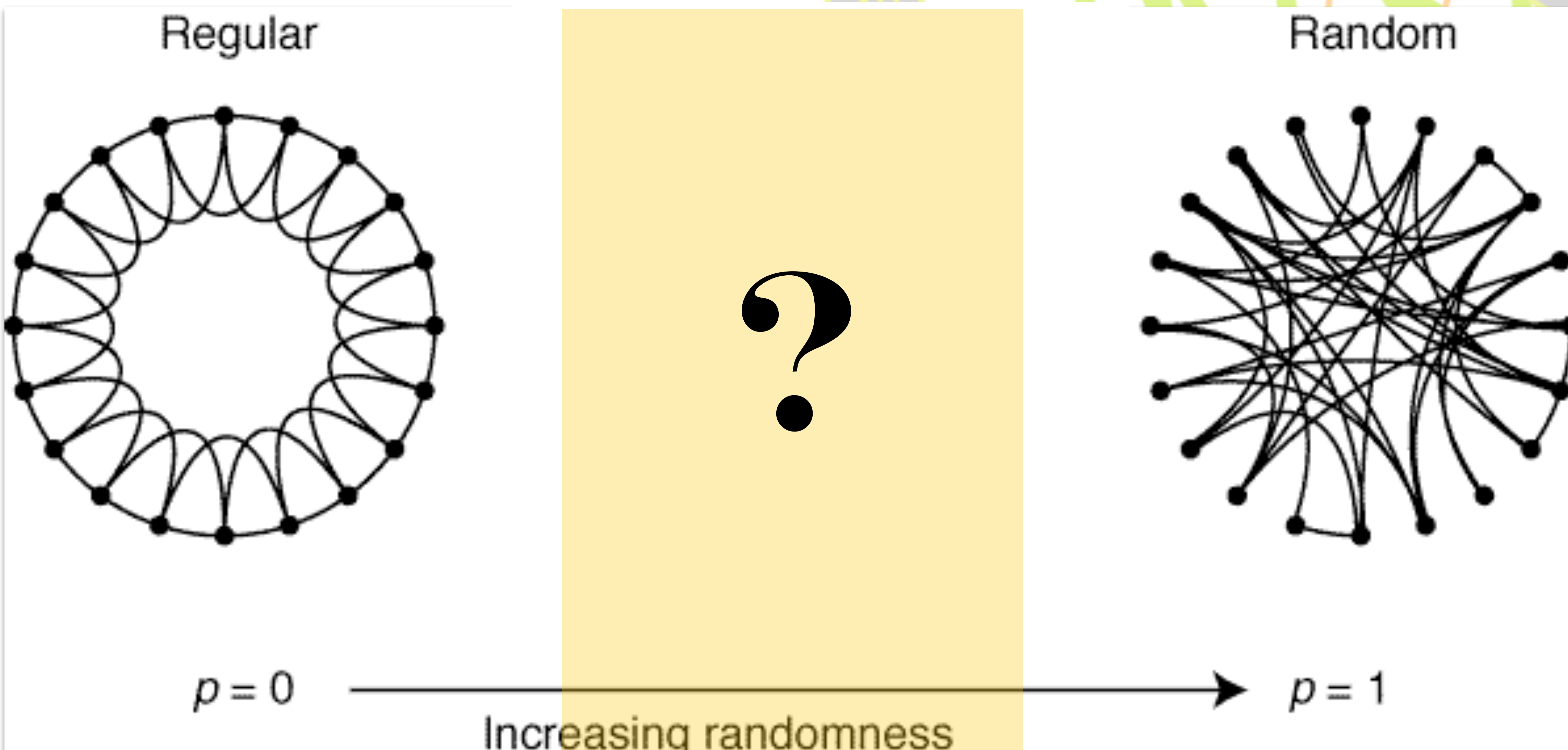
$$L \simeq \frac{\ln N}{\ln\langle k\rangle} \qquad\qquad C = \frac{\langle k\rangle}{N}$$



Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

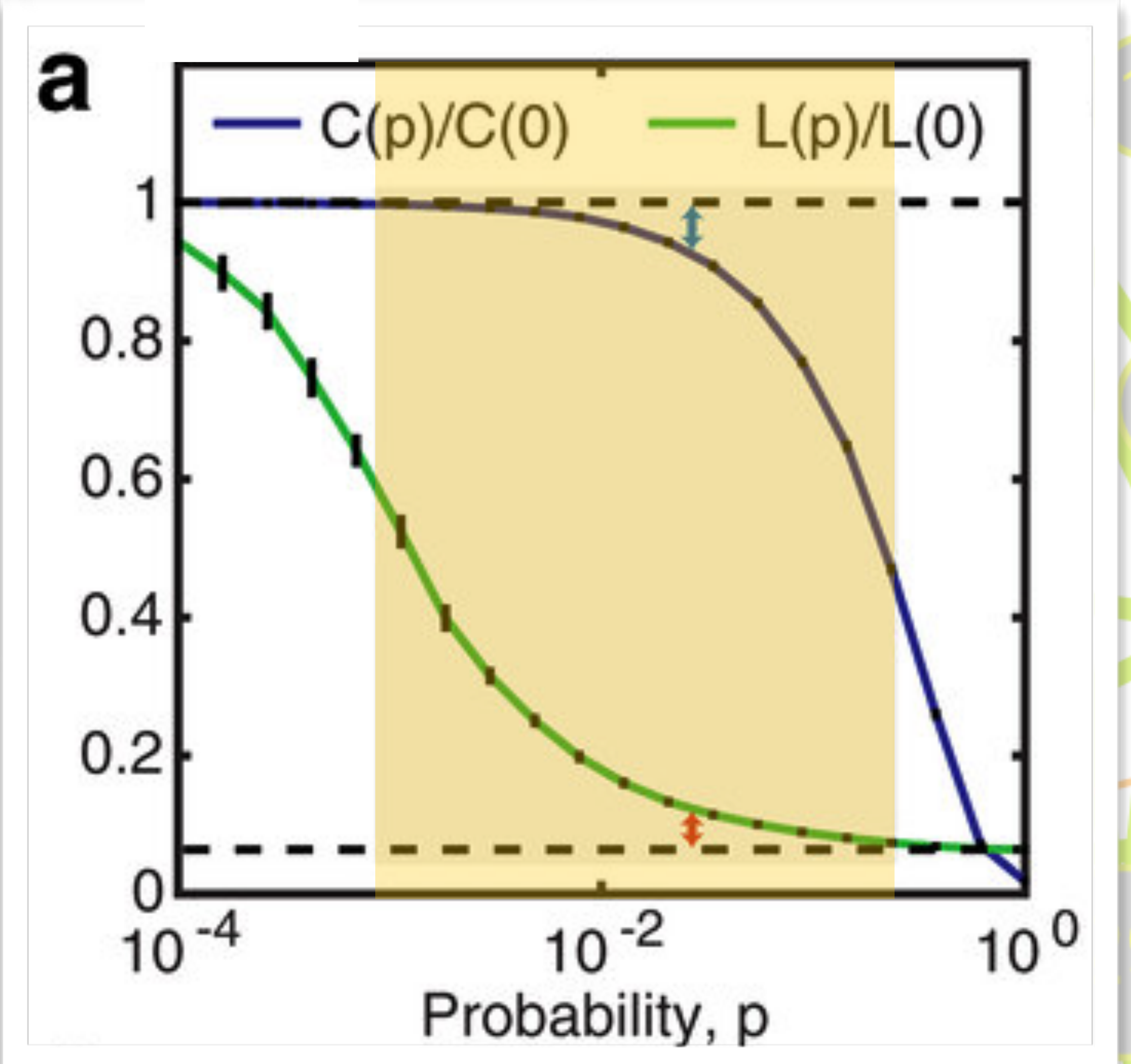
$L = $ **?** $\qquad\qquad C =$ **?**



@gomezgardenes

Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

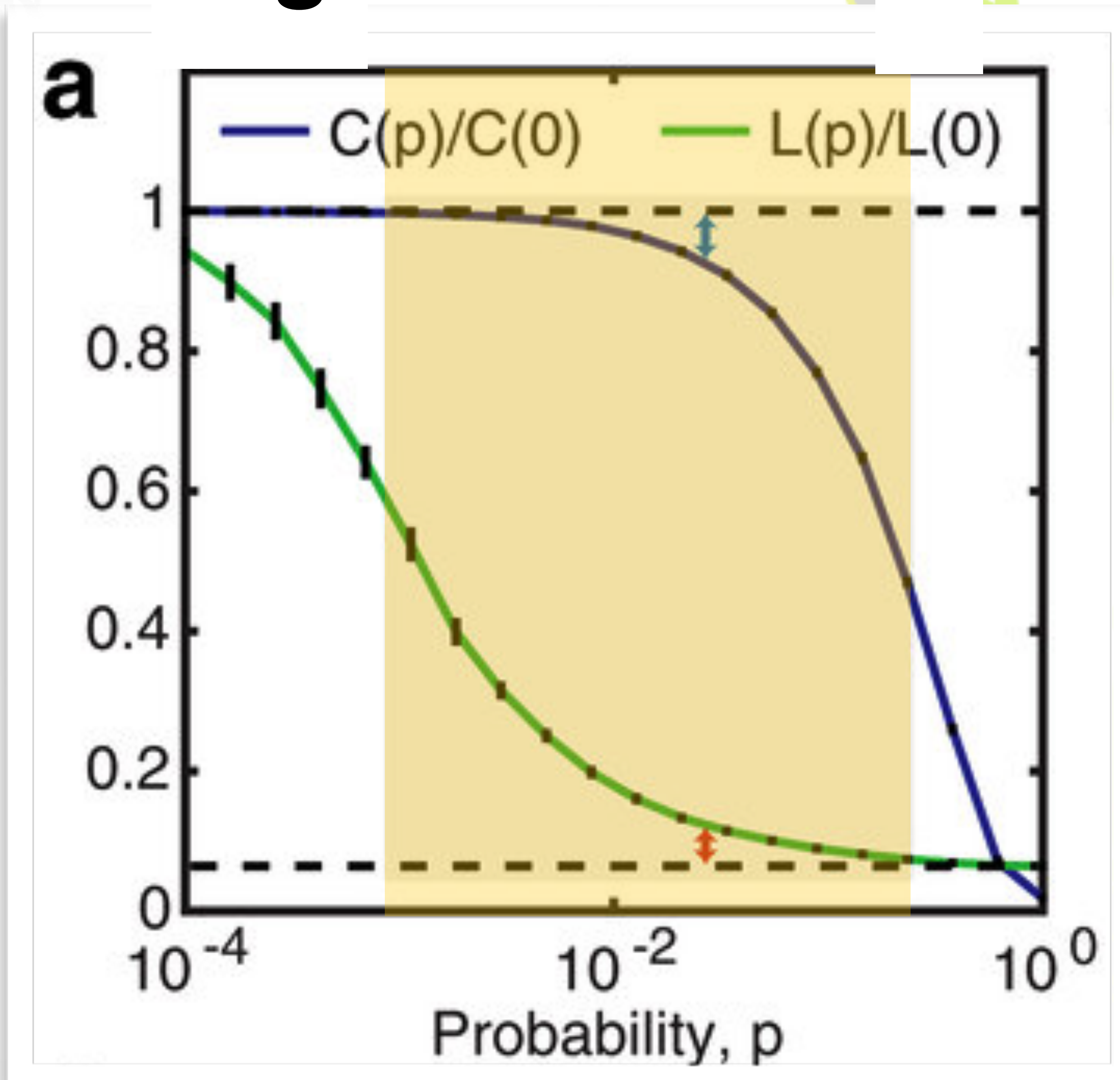


@gomezgardenes

Watts-Strogatz Model

D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

Degree distribution?

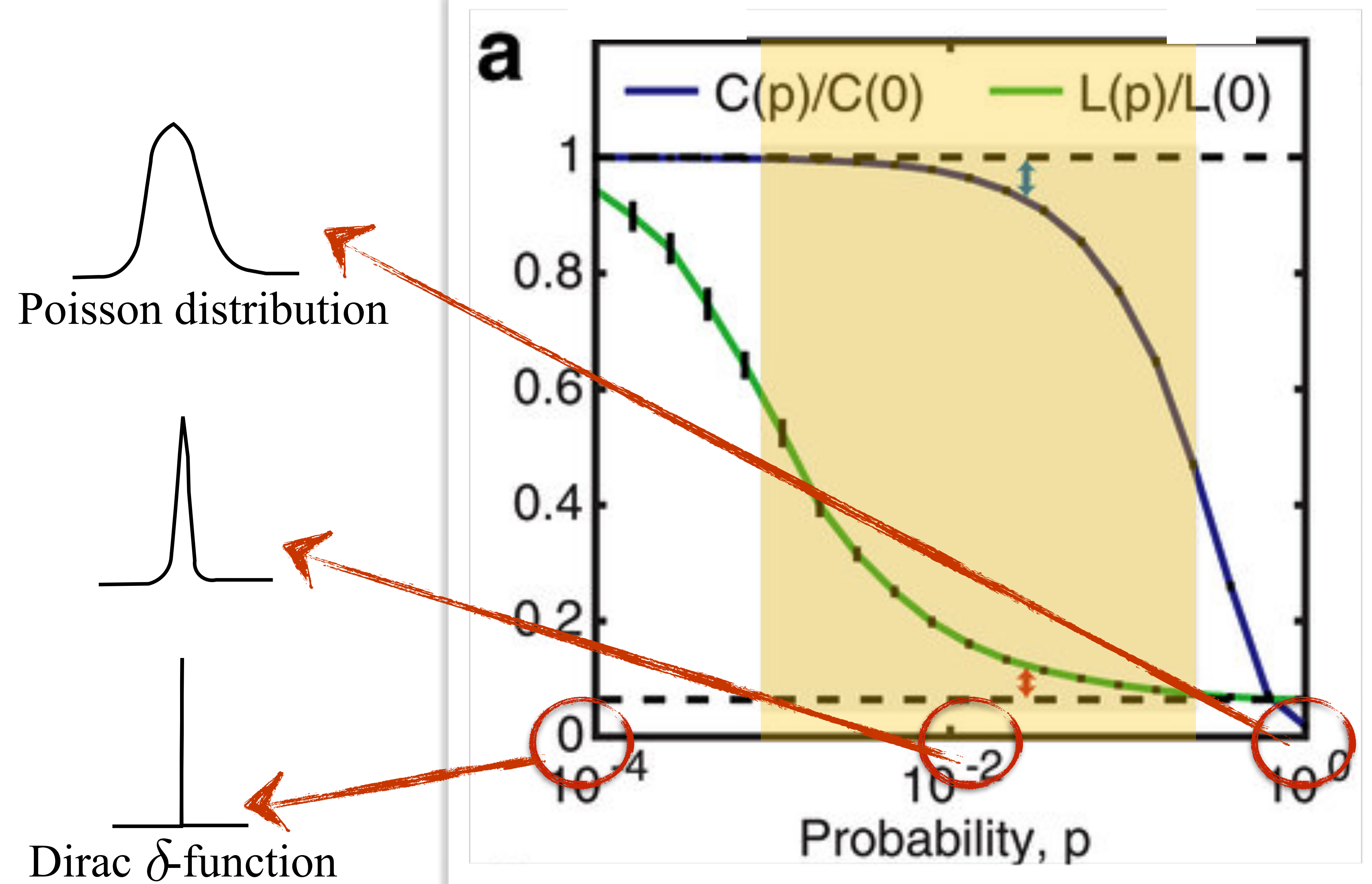


@gomezgardenes

Watts-Strogatz Model

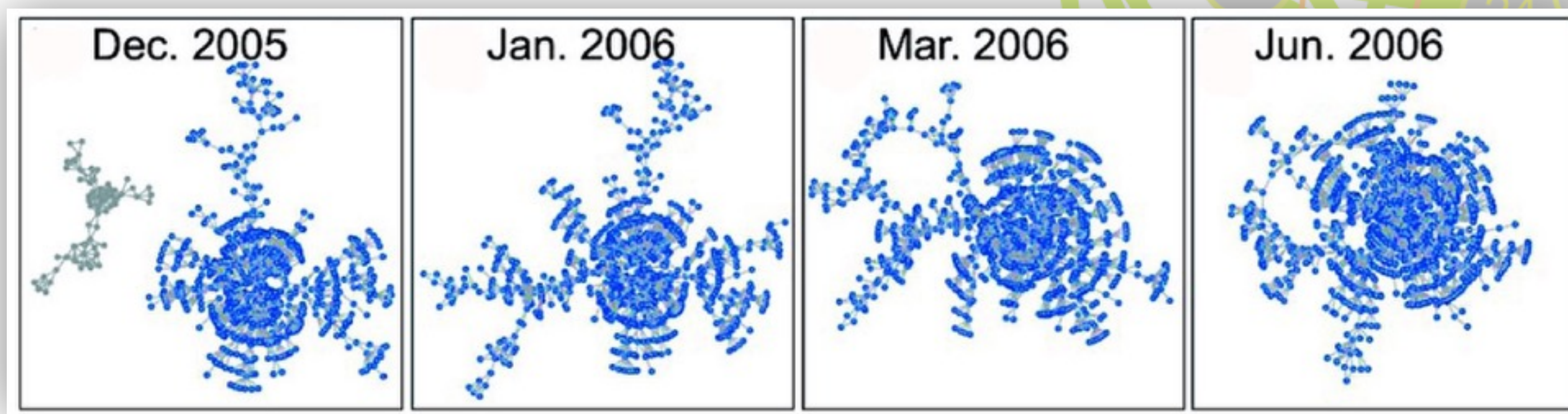
D. WATTS & S.H. STROGATZ, NATURE 393, 440 (1998)

Degree distribution?



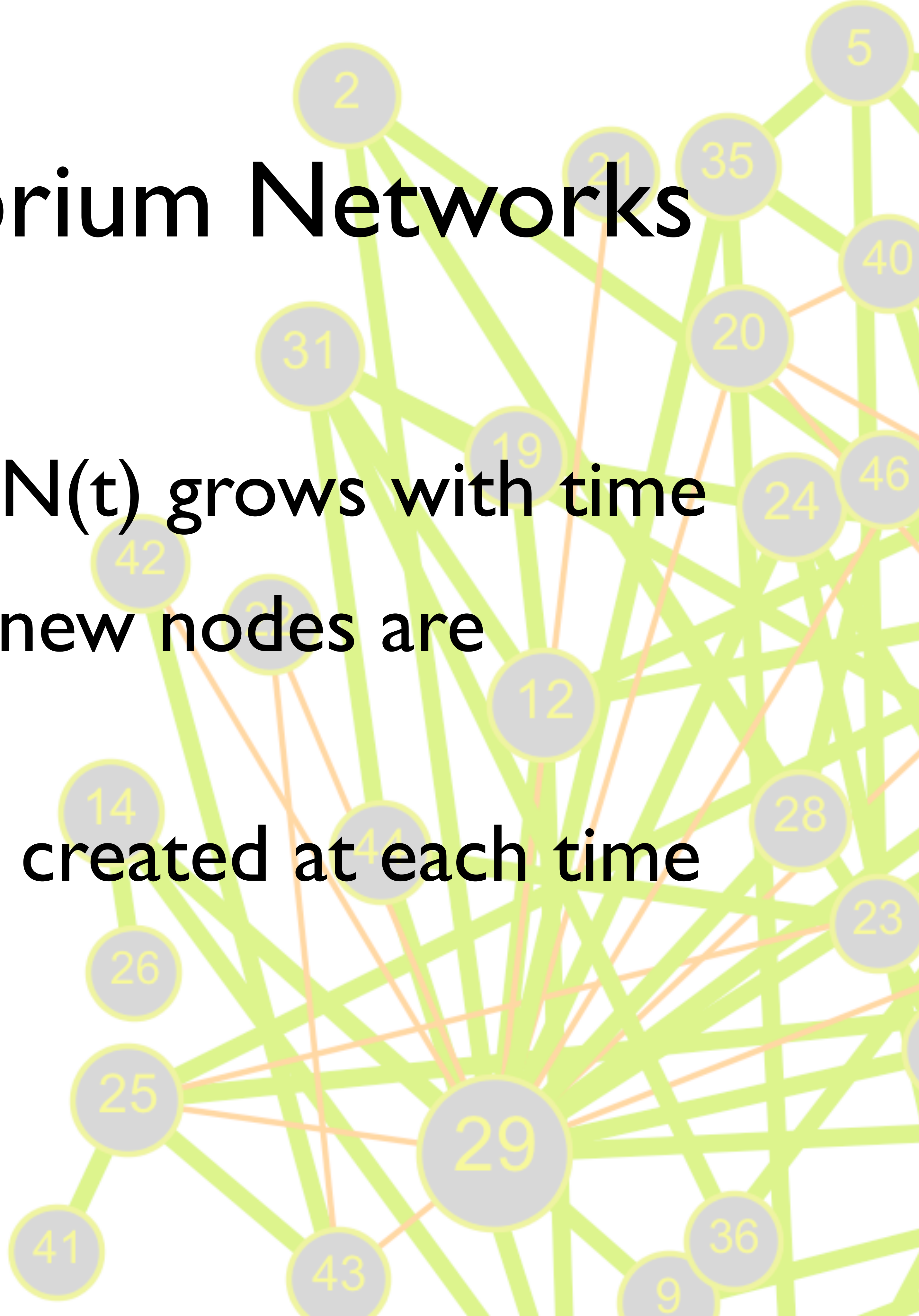
@gomezgardenes

… but Complex Networks evolve!



Non-Equilibrium Networks

- Number of nodes N(t) grows with time
- At each time step new nodes are incorporated
- New links are also created at each time step

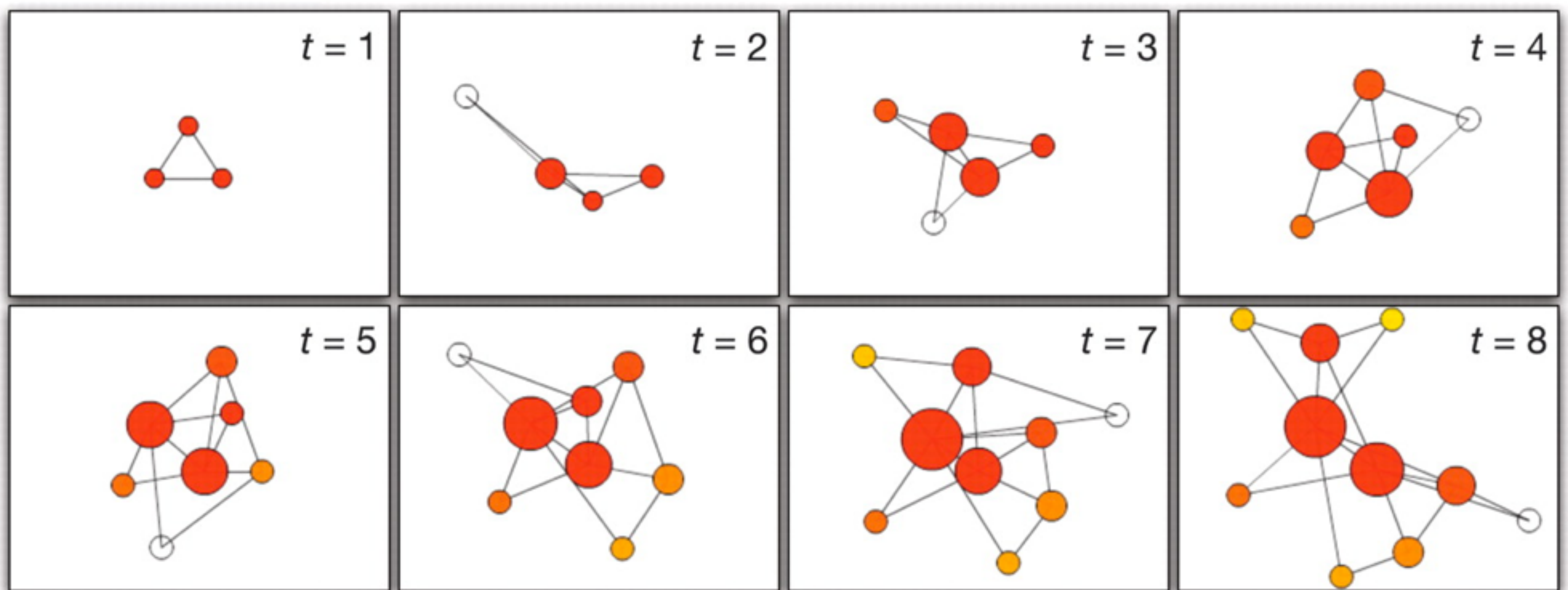


@gomezgardenes

Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

- At each time step 1 new node is incorporated
- The new node launches m new links to the already existing nodes



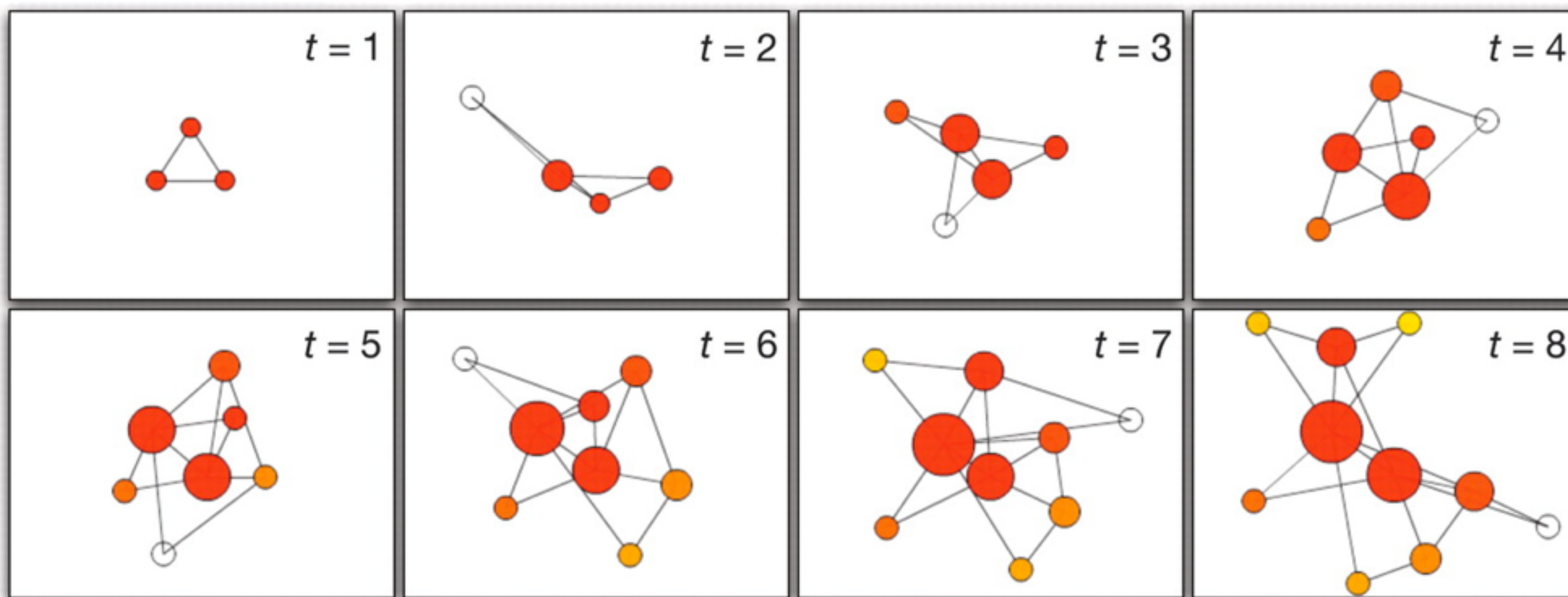
Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

- The probability that a node i receives a link from the newcomer is:

$$\Pi_i(t) = \frac{k_i(t)}{\sum_{j=1}^{t+m_0-1} k_j(t)}$$

(Preferential attachment rule)



Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

- The time evolution of the degree of a node is given by:

$$\frac{\partial k_i}{\partial t} = m \frac{k_i(t)}{\sum_{j=1}^{t+m_0-1} k_j(t)} \quad \textbf{with} \quad k_i(t=t_i) = m$$

- …whose solution is:

@gomezgardenes

Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

- The time evolution of the degree of a node is given by:

$$\frac{\partial k_i}{\partial t} = m \frac{k_i(t)}{\sum_{j=1}^{t+m_0-1} k_j(t)} \qquad \text{with} \quad k_i(t = t_i) = m$$

- …whose solution is:

$$k_i(t) = 0 \qquad t < t_i$$

$$k_i(t) = m \left(\frac{t}{t_i}\right)^{1/2} \qquad t > t_i$$

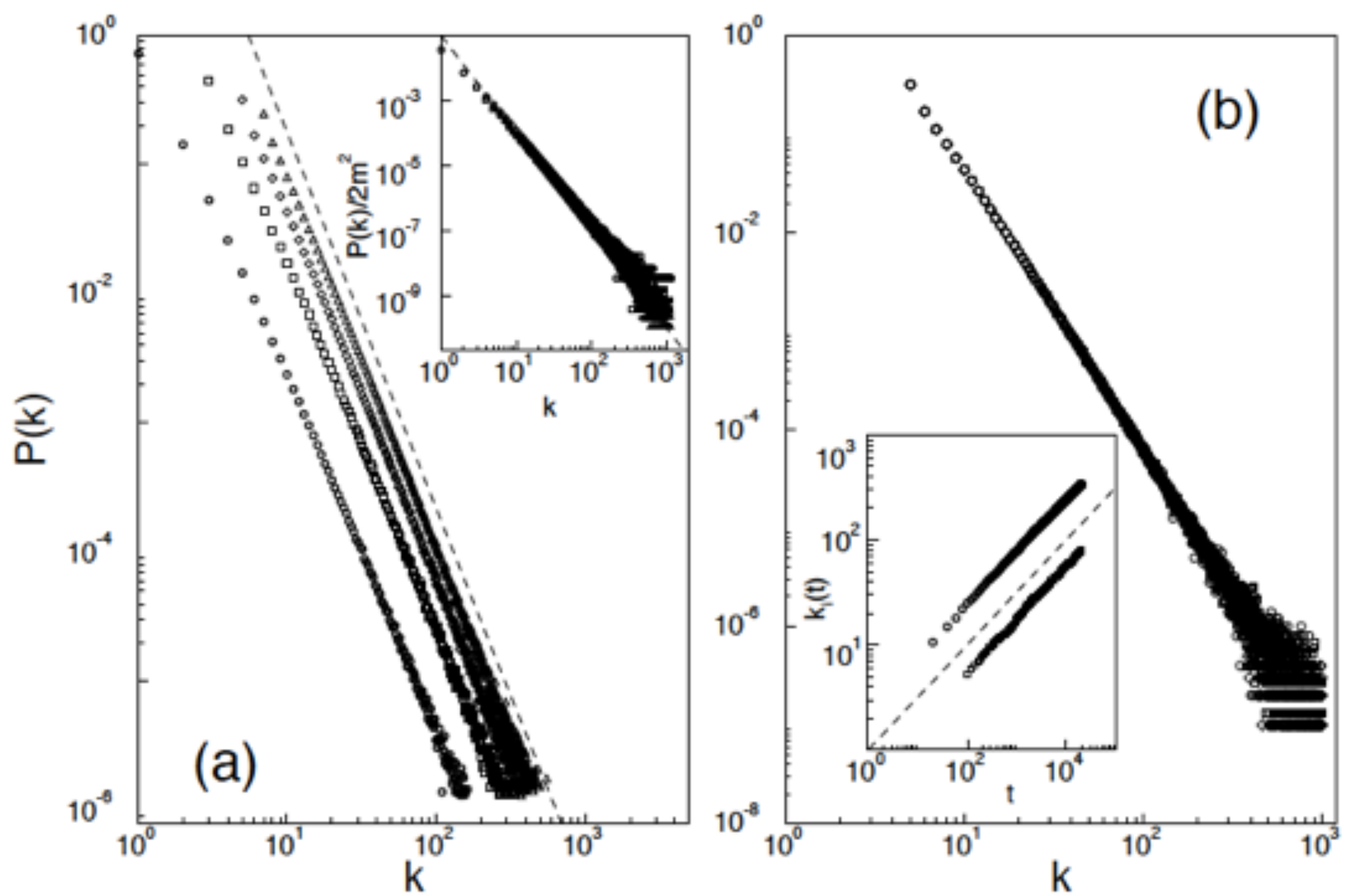
@gomezgardenes

Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

- …that finally yields:

$$P(k) = \frac{2m(m+1)}{k(K+1)(k+2)} \simeq k^{-3}$$

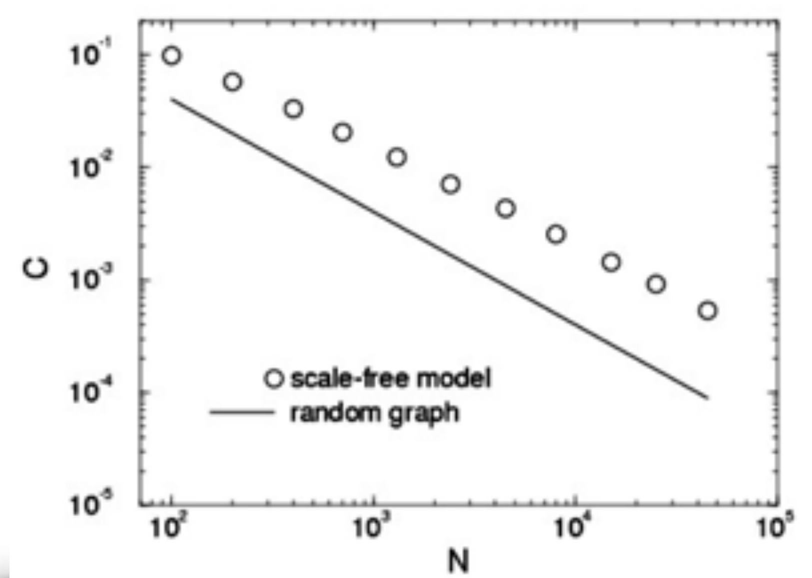


@gomezgardenes

Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

Clustering Coefficient



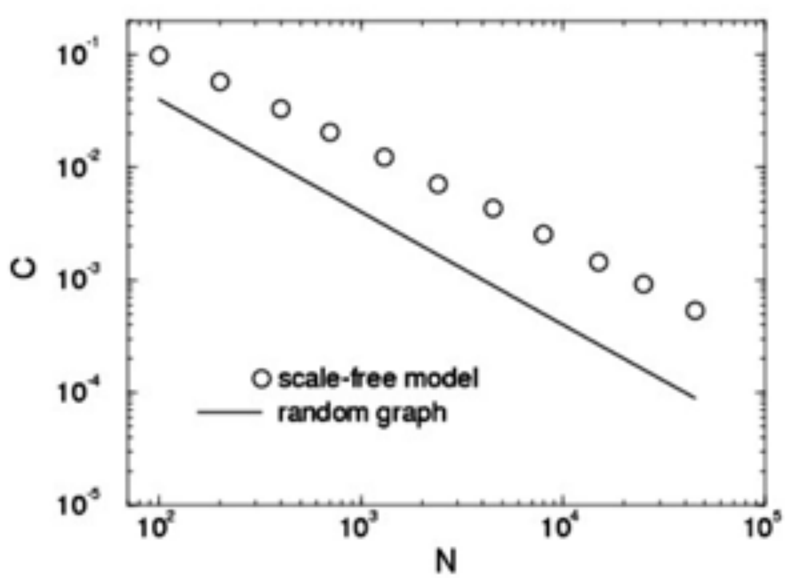
$$C_{BA} \sim N^{-0.75}$$

$$C_{ER} \sim N^{-1}$$

Barabàsi-Albert Model

A.L. BARABÀSI & R. ALBERT, SCIENCE 286, 509 (1999)

Clustering Coefficient



$$C_{BA} \sim N^{-0.75}$$

$$C_{ER} \sim N^{-1}$$

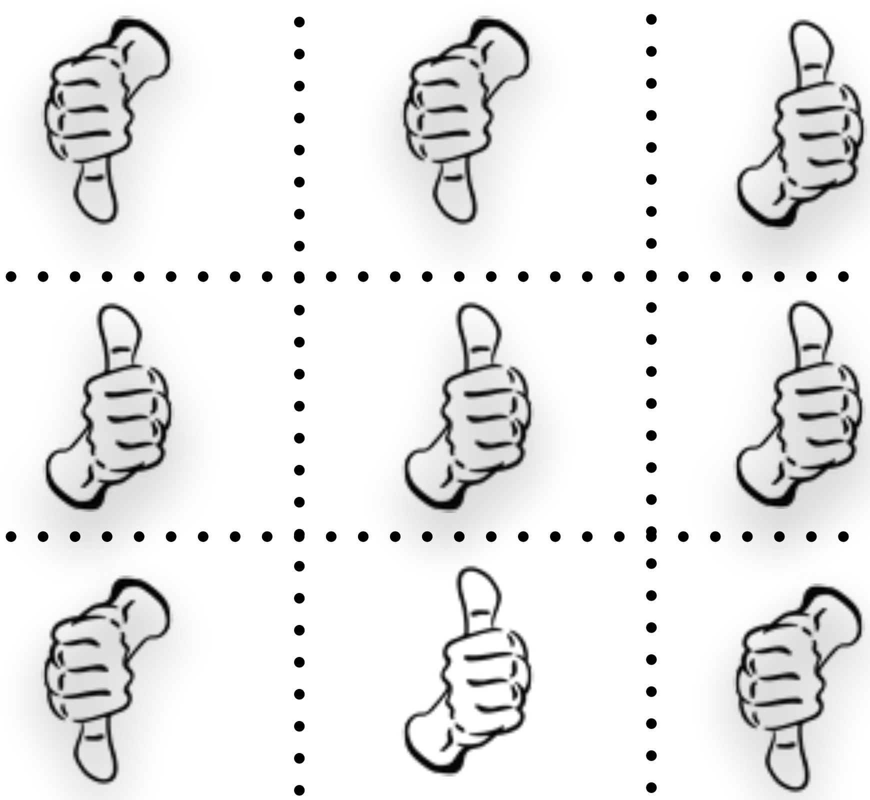
Average Path length

$$L_{BA} \sim \frac{\ln N}{\ln \ln N}$$

$$L_{ER} \sim \ln N$$

Overview

	ER	WS	BA
• Degree distribution			
• Average length			
• Clustering Coefficient			



A number of variations of the former network models available

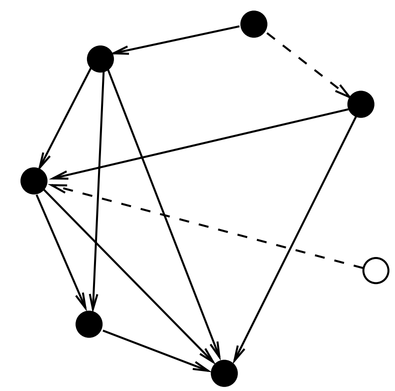
@gomezgardenes

Alternative Formulations

DOROGOVTSEV-MENDES-SAMUKHIN
PHYS. REV. LETT. 85, 4633 (2000)

Scale-free with tunable γ

$$\Pi_i(t) = \frac{k_i + \alpha}{\sum_{j=1}^{m_0+t-1}(k_j(t)+\alpha)}$$



$$\alpha \in (-m, \infty)$$

@gomezgardenes

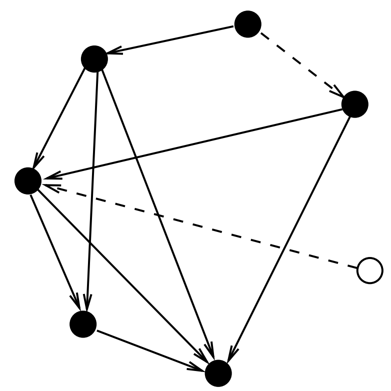
Alternative Formulations

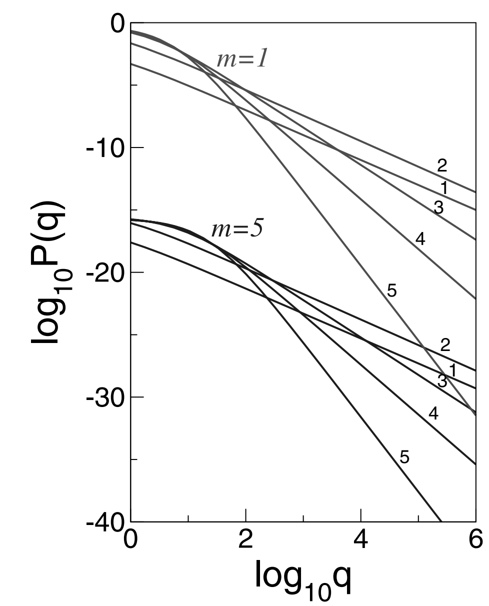
Dorogovtsev-Mendes-Samukhin
Phys. Rev. Lett. 85, 4633 (2000)

Scale-free with tunable γ

$$\Pi_i(t) = \frac{k_i + \alpha}{\sum_{j=1}^{m_0+t-1}(k_j(t)+\alpha)}$$

$$\alpha \in (-m, \infty) \longrightarrow \gamma = 3 + \frac{\alpha}{m}$$





@gomezgardenes

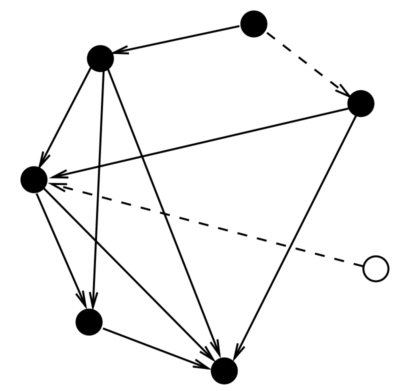
Alternative Formulations

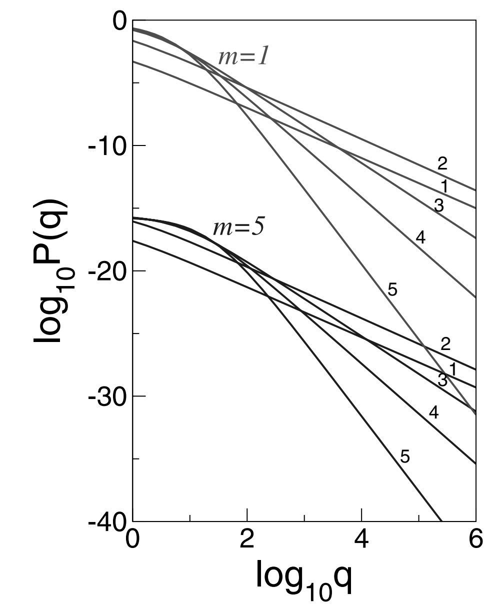
Scale-free with tunable γ

DOROGOVTSEV-MENDES-SAMUKHIN
PHYS. REV. LETT. 85, 4633 (2000)

$$\Pi_i(t) = \frac{k_i + \alpha}{\sum_{j=1}^{m_0+t-1}(k_j(t)+\alpha)}$$

$$\alpha \in (-m, \infty) \longrightarrow \gamma = 3 + \frac{\alpha}{m}$$





Scale-free with high clustering

HOLME-KIM
PHYS. REV. E 65, 026107 (2002)

- First link: follow usual PA rule
- For each of the $m-1$ links:

  (i) With probability $(1-q)$: usual PA

  (ii) With probability q: Attach to one neighbor of the first chosen node

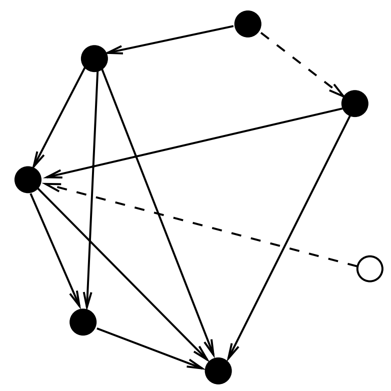
@gomezgardenes

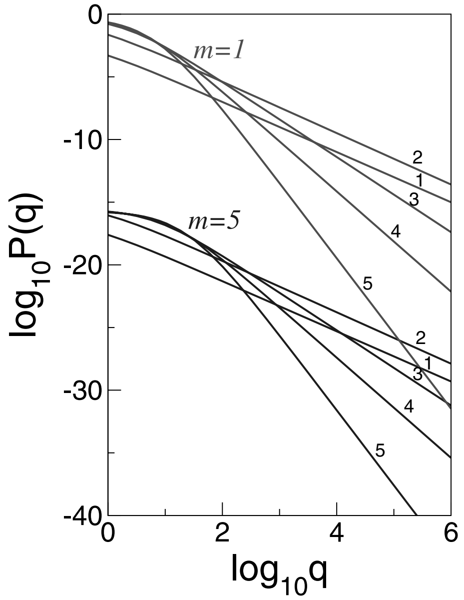
Alternative Formulations

Scale-free with tunable γ

DOROGOVTSEV-MENDES-SAMUKHIN
PHYS. REV. LETT. 85, 4633 (2000)

$$\Pi_i(t) = \frac{k_i + \alpha}{\sum_{j=1}^{m_0+t-1}(k_j(t)+\alpha)}$$





$\alpha \in (-m, \infty) \longrightarrow \gamma = 3 + \frac{\alpha}{m}$

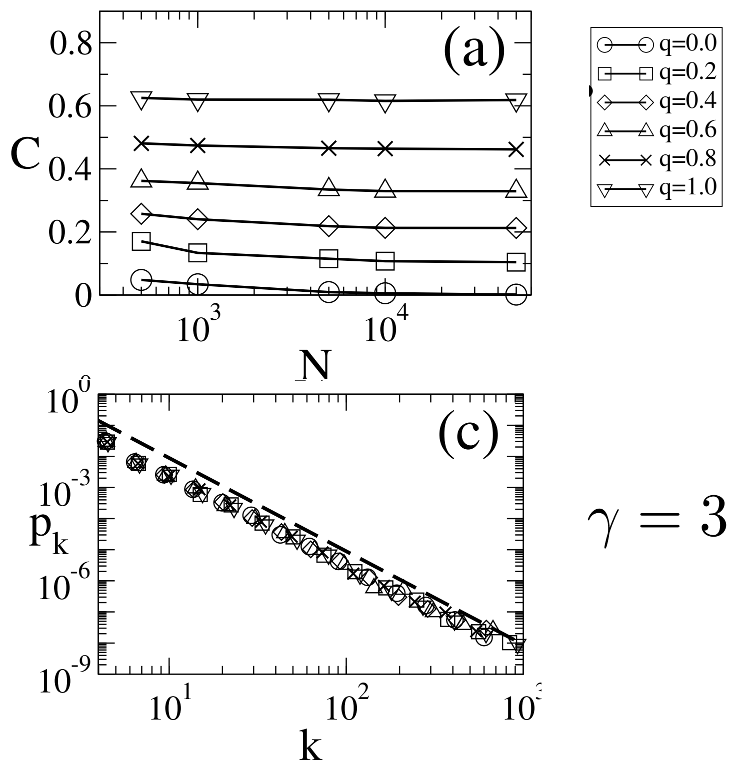
Scale-free with high clustering

HOLME-KIM
PHYS. REV. E 65, 026107 (2002)

- First link: follow usual PA rule
- For each of the $m-1$ links:

  (i) With probability $(1-q)$: usual PA

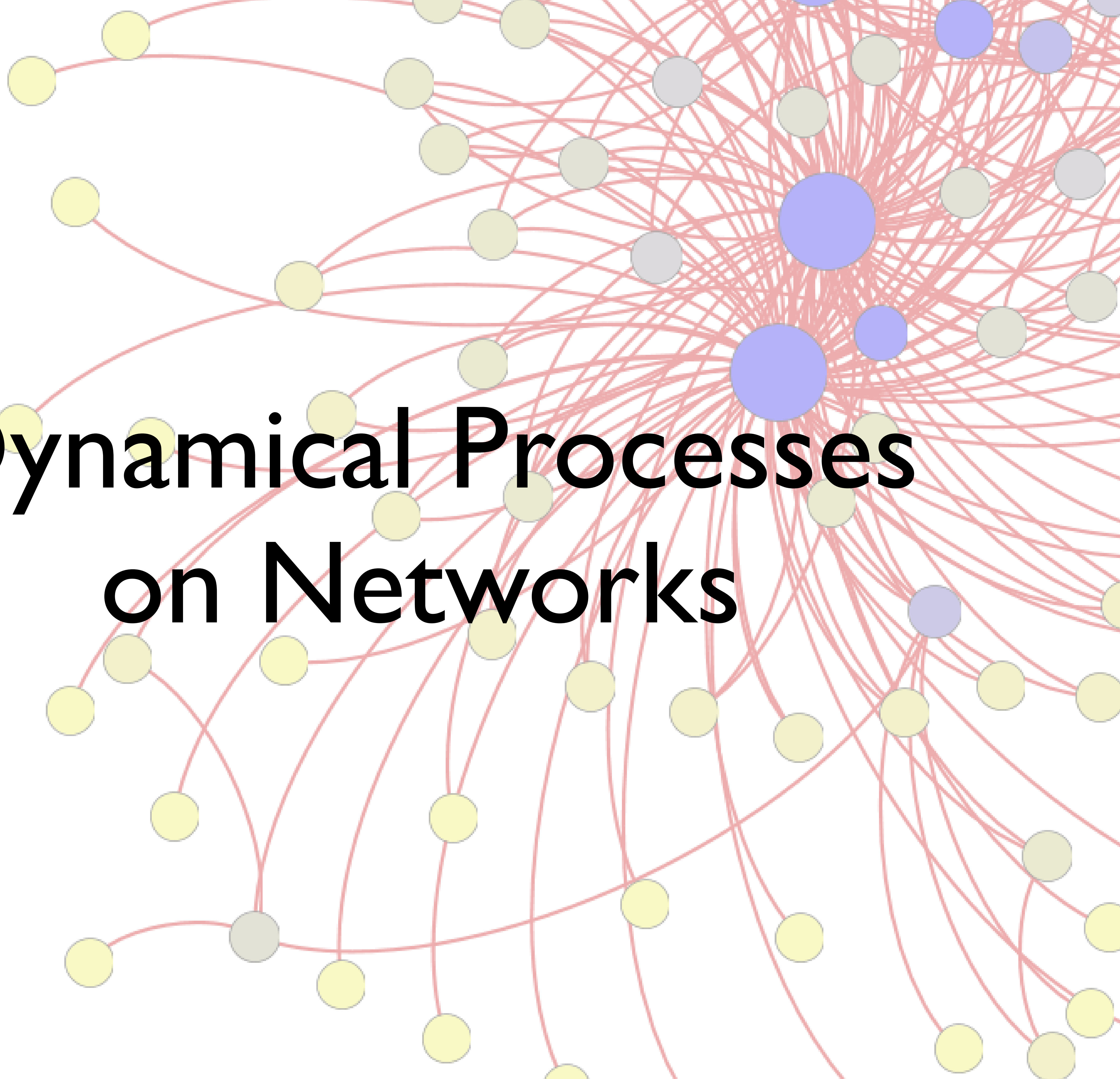
  (ii) With probability q: Attach to one neighbor of the first chosen node



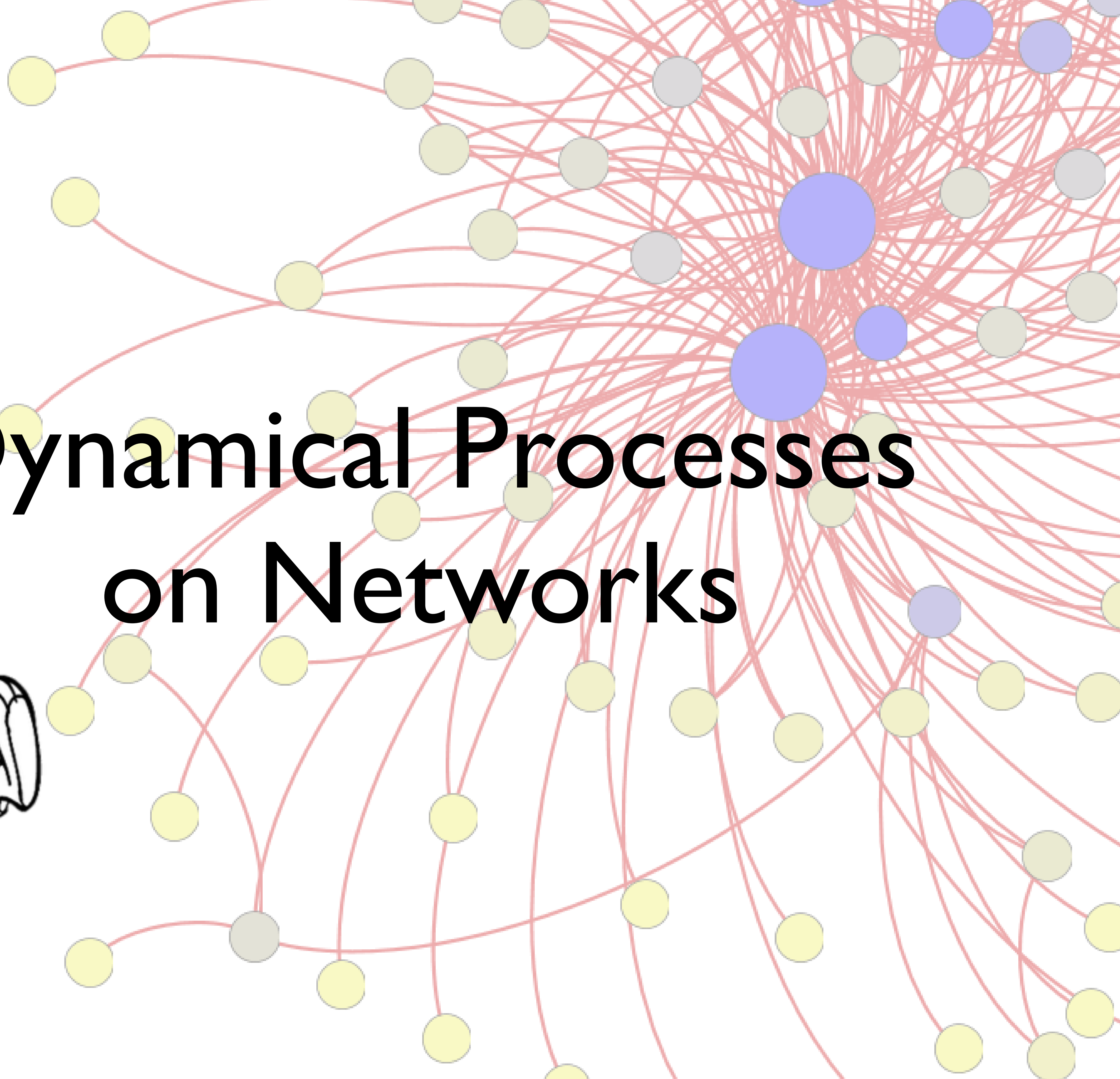
$\gamma = 3$

@gomezgardenes

Dynamical Processes on Networks



@gomezgardenes



Dynamical Processes on Networks

@gomezgardenes

OPEN ACCESS Freely available online PLOS ONE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

Delia Mocanu[1], **Andrea Baronchelli**[1], **Nicola Perra**[1], **Bruno Gonçalves**[2], **Qian Zhang**[1], **Alessandro Vespignani**[1,3,4]*

1 Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, Massachusetts, United States of America, **2** Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France, **3** Institute for Quantitative Social Sciences at Harvard University, Cambridge, Massachusetts, United States of America, **4** Institute for Scientific Interchange Foundation, Turin, Italy

Abstract

Large scale analysis and statistics of socio-technical systems that just a few short years ago would have required the use of consistent economic and human resources can nowadays be conveniently performed by mining the enormous amount of digital data produced by human activities. Although a characterization of several aspects of our societies is emerging from the data revolution, a number of questions concerning the reliability and the biases inherent to the big data "proxies" of social life are still open. Here, we survey worldwide linguistic indicators and trends through the analysis of a large-scale dataset of microblogging posts. We show that available data allow for the study of language geography at scales ranging from country-level aggregation to specific city neighborhoods. The high resolution and coverage of the data allows us to investigate different indicators such as the linguistic homogeneity of different countries, the touristic seasonal patterns within countries and the geographical distribution of different languages in multilingual regions. This work highlights the potential of geolocalized studies of open data sources to improve current analysis and develop indicators for major social phenomena in specific communities.

Citation: Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981

Editor: Yamir Moreno, University of Zaragoza, Spain

Received January 5, 2013; **Accepted** March 18, 2013; **Published** April 18, 2013

Funding: The authors acknowledge the support by the National Science Foundation ICES award CCF-1101743. For the analysis of data data outside of the United States of America the authors acknowledge the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

nature Vol 457 | 19 February 2009 | **doi:10.1038/nature07634**

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

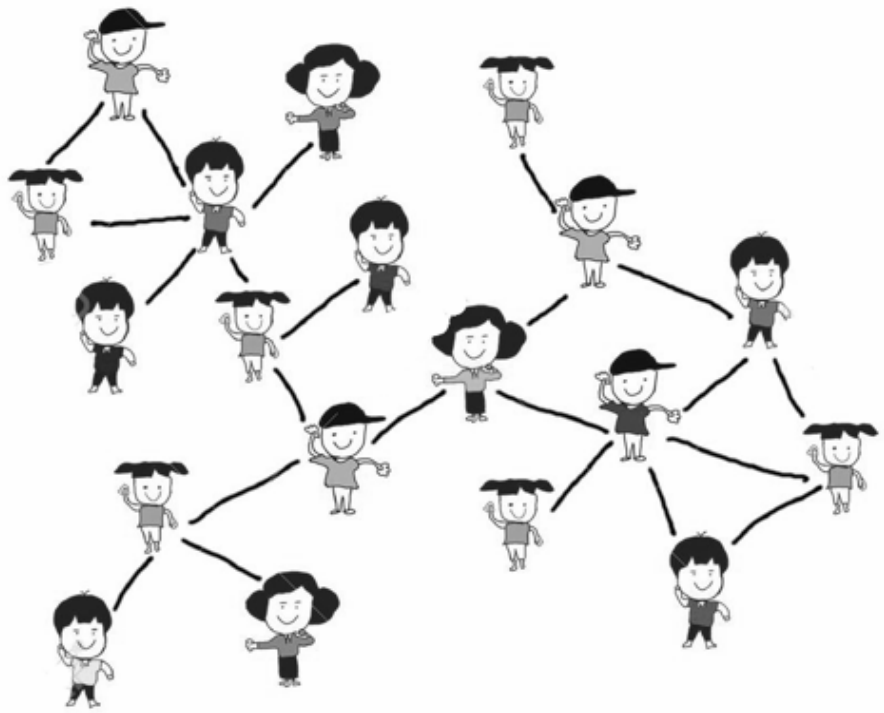
Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\mathrm{logit}(I(t)) = \alpha\,\mathrm{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\mathrm{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza

@gomezgardenes



OPEN ACCESS Freely available online PLOS ONE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

Delia Mocanu[1], Andrea Baronchelli[1], Nicola Perra[1], Bruno Gonçalves[2], Qian Zhang[1], Alessandro Vespignani[1,3,4]*

1 Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, Massachusetts, United States of America, 2 Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France, 3 Institute for Quantitative Social Sciences at Harvard University, Cambridge, Massachusetts, United States of America, 4 Institute for Scientific Interchange Foundation, Turin, Italy

Abstract

Large scale analysis and statistics of socio-technical systems that just a few short years ago would have required the use of consistent economic and human resources can nowadays be conveniently performed by mining the enormous amount of digital data produced by human activities. Although a characterization of several aspects of our societies is emerging from the data revolution, a number of questions concerning the reliability and the biases inherent to the big data "proxies" of social life are still open. Here, we survey worldwide linguistic indicators and trends through the analysis of a large-scale dataset of microblogging posts. We show that available data allow for the study of language geography at scales ranging from country-level aggregation to specific city neighborhoods. The high resolution and coverage of the data allows us to investigate different indicators such as the linguistic homogeneity of different countries, the touristic seasonal patterns within countries and the geographical distribution of different languages in multilingual regions. This work highlights the potential of geolocalized studies of open data sources to improve current analysis and develop indicators for major social phenomena in specific communities.

Citation: Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981

Editor: Yamir Moreno, University of Zaragoza, Spain

Received January 5, 2013; **Accepted** March 18, 2013; **Published** April 18, 2013

Funding: The authors acknowledge the support by the National Science Foundation ICES award CCF-1101743. For the analysis of data data outside of the United States of America the authors acknowledge the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

nature Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

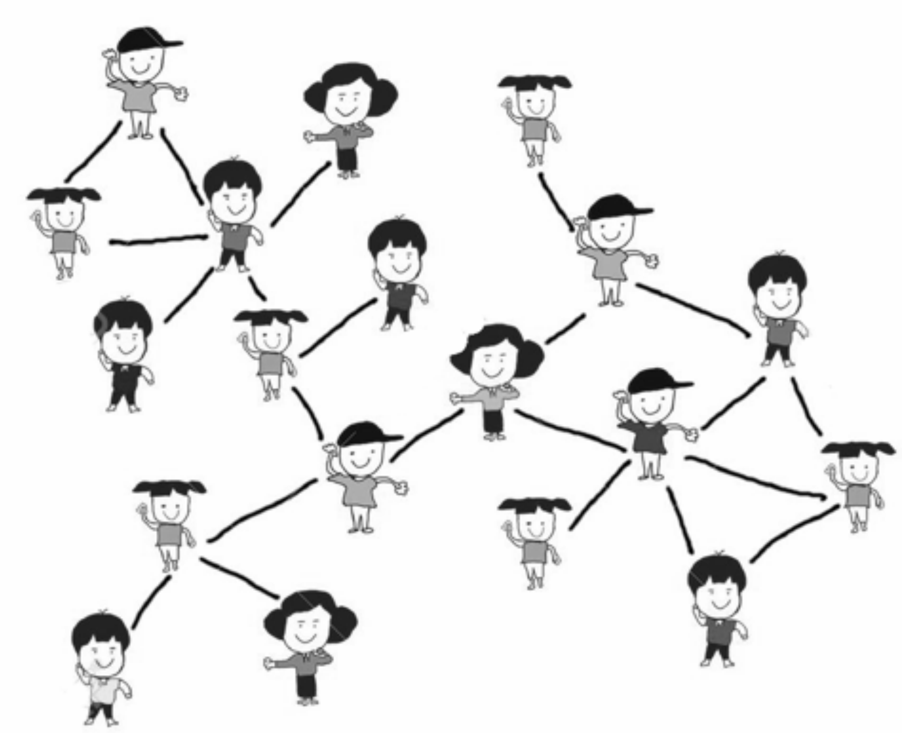
Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

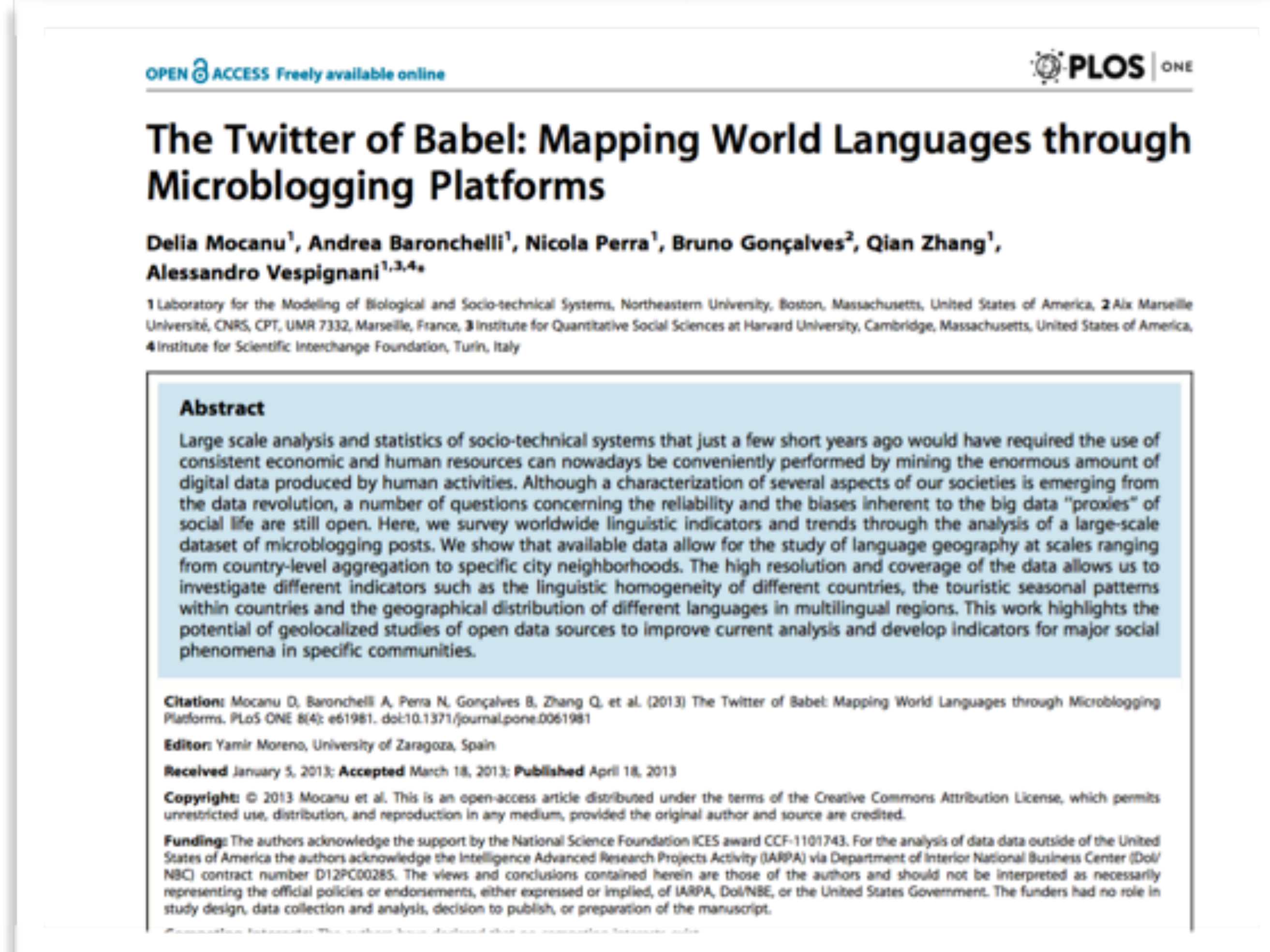
By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\text{logit}(I(t)) = \alpha\text{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\text{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza

@gomezgardenes





OPEN ACCESS Freely available online | PLOS ONE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

Delia Mocanu[1], Andrea Baronchelli[1], Nicola Perra[1], Bruno Gonçalves[2], Qian Zhang[1], Alessandro Vespignani[1,3,4]*

1 Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, Massachusetts, United States of America, **2** Aix Marseille Université, CNRS, CPT, UMR 7332, Marseille, France, **3** Institute for Quantitative Social Sciences at Harvard University, Cambridge, Massachusetts, United States of America, **4** Institute for Scientific Interchange Foundation, Turin, Italy

Abstract

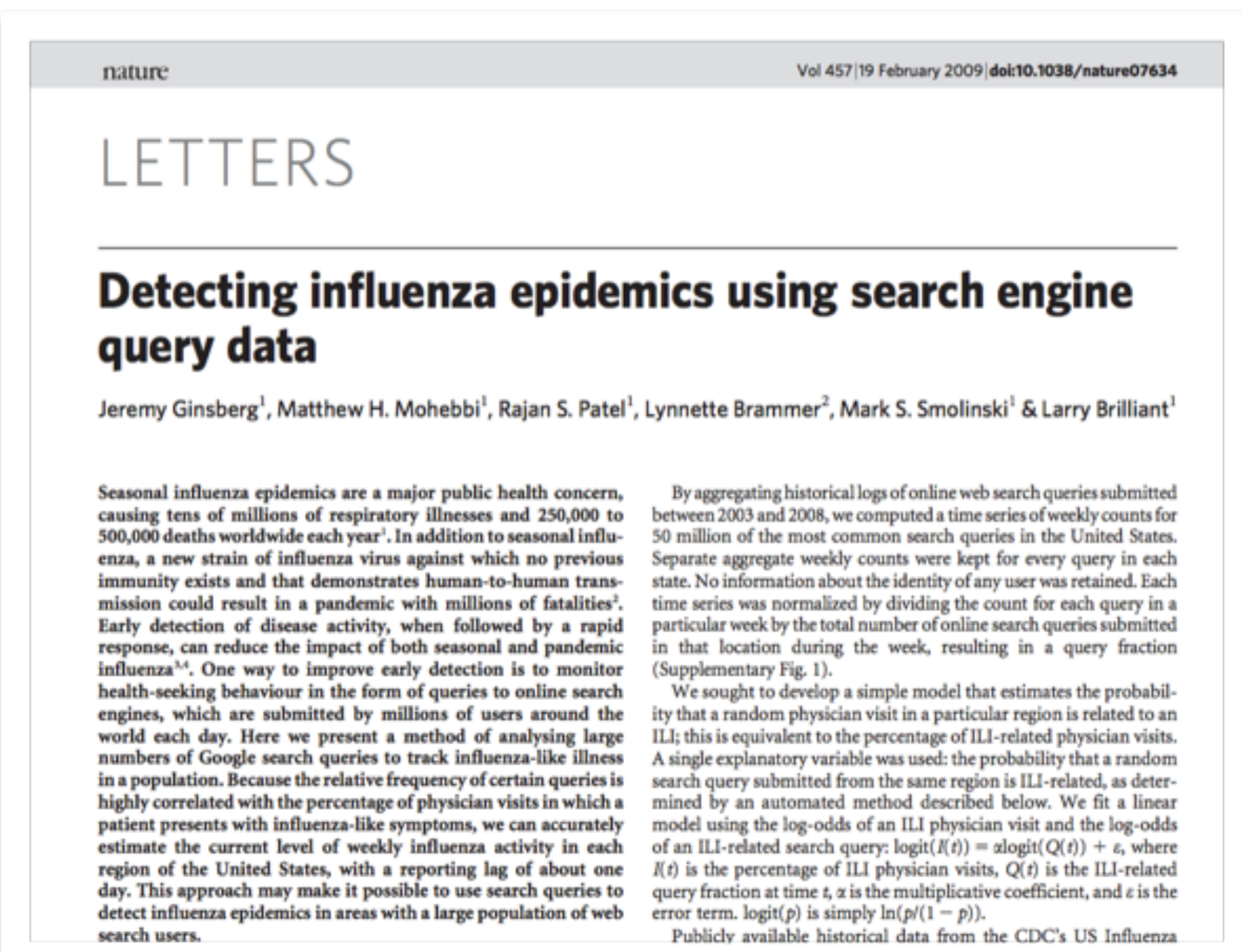
Large scale analysis and statistics of socio-technical systems that just a few short years ago would have required the use of consistent economic and human resources can nowadays be conveniently performed by mining the enormous amount of digital data produced by human activities. Although a characterization of several aspects of our societies is emerging from the data revolution, a number of questions concerning the reliability and the biases inherent to the big data "proxies" of social life are still open. Here, we survey worldwide linguistic indicators and trends through the analysis of a large-scale dataset of microblogging posts. We show that available data allow for the study of language geography at scales ranging from country-level aggregation to specific city neighborhoods. The high resolution and coverage of the data allows us to investigate different indicators such as the linguistic homogeneity of different countries, the touristic seasonal patterns within countries and the geographical distribution of different languages in multilingual regions. This work highlights the potential of geolocalized studies of open data sources to improve current analysis and develop indicators for major social phenomena in specific communities.

Citation: Mocanu D, Baronchelli A, Perra N, Gonçalves B, Zhang Q, et al. (2013) The Twitter of Babel: Mapping World Languages through Microblogging Platforms. PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981

Editor: Yamir Moreno, University of Zaragoza, Spain

Received January 5, 2013; **Accepted** March 18, 2013; **Published** April 18, 2013

Funding: The authors acknowledge the support by the National Science Foundation ICES award CCF-1101743. For the analysis of data data outside of the United States of America the authors acknowledge the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.



nature | Vol 457 | 19 February 2009 | **doi:10.1038/nature07634**

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\mathrm{logit}(I(t)) = \alpha\,\mathrm{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\mathrm{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza

SO WHAT?

@gomezgardenes

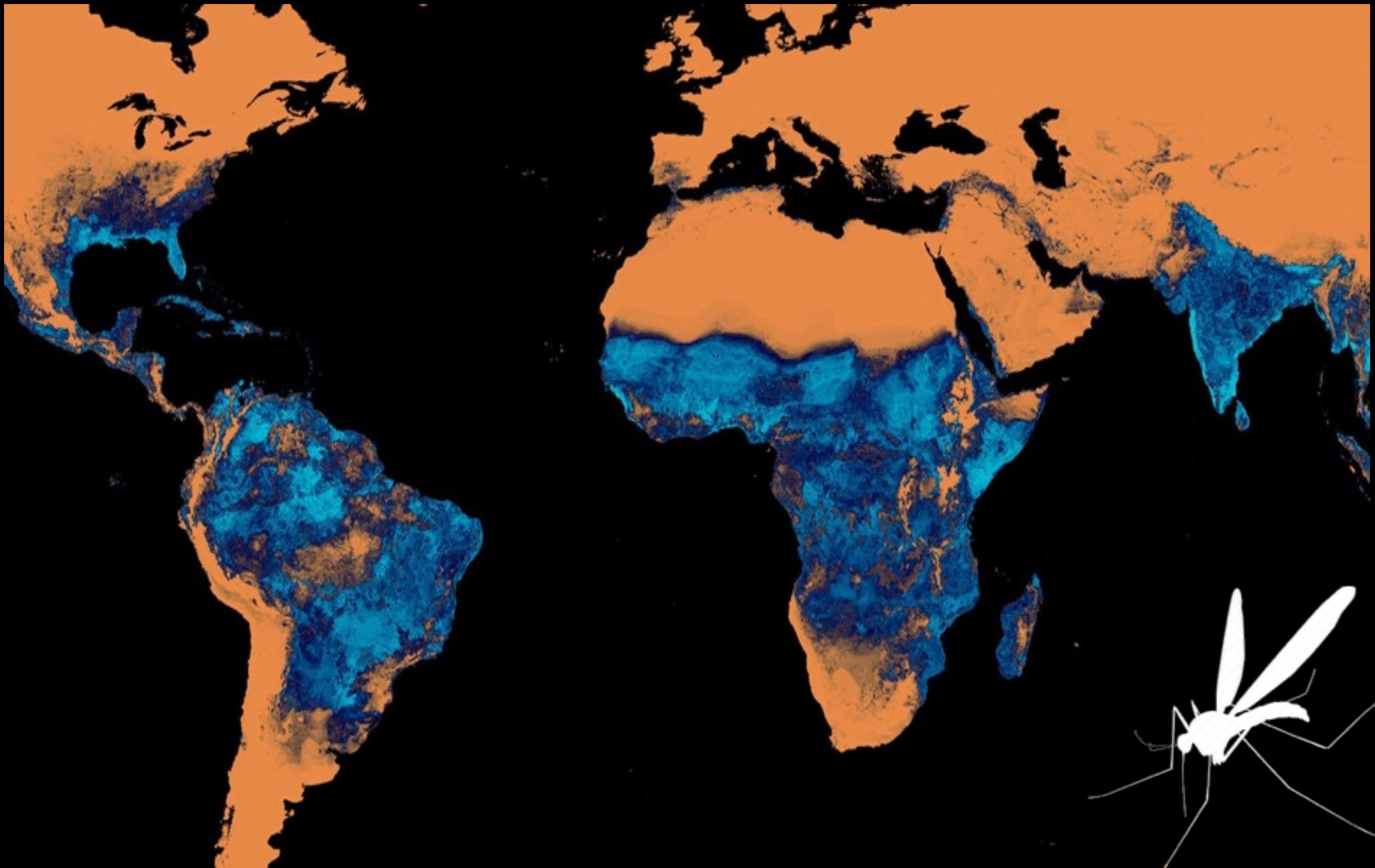
Fundamental Question

@gomezgardenes

Fundamental Question

?

Fundamental Question



@gomezgardenes

Fundamental Question

@gomezgardenes

Main Dynamical Processes

- Simple diffusion processes
- Cascades (Failures and Attacks)
- Contagion processes
- Diffusion with queues
- Synchronization
- Evolutionary games
- Chaotic dynamics
- ...

@gomezgardenes

REVIEWS OF MODERN PHYSICS, VOLUME 80, OCTOBER–DECEMBER 2008

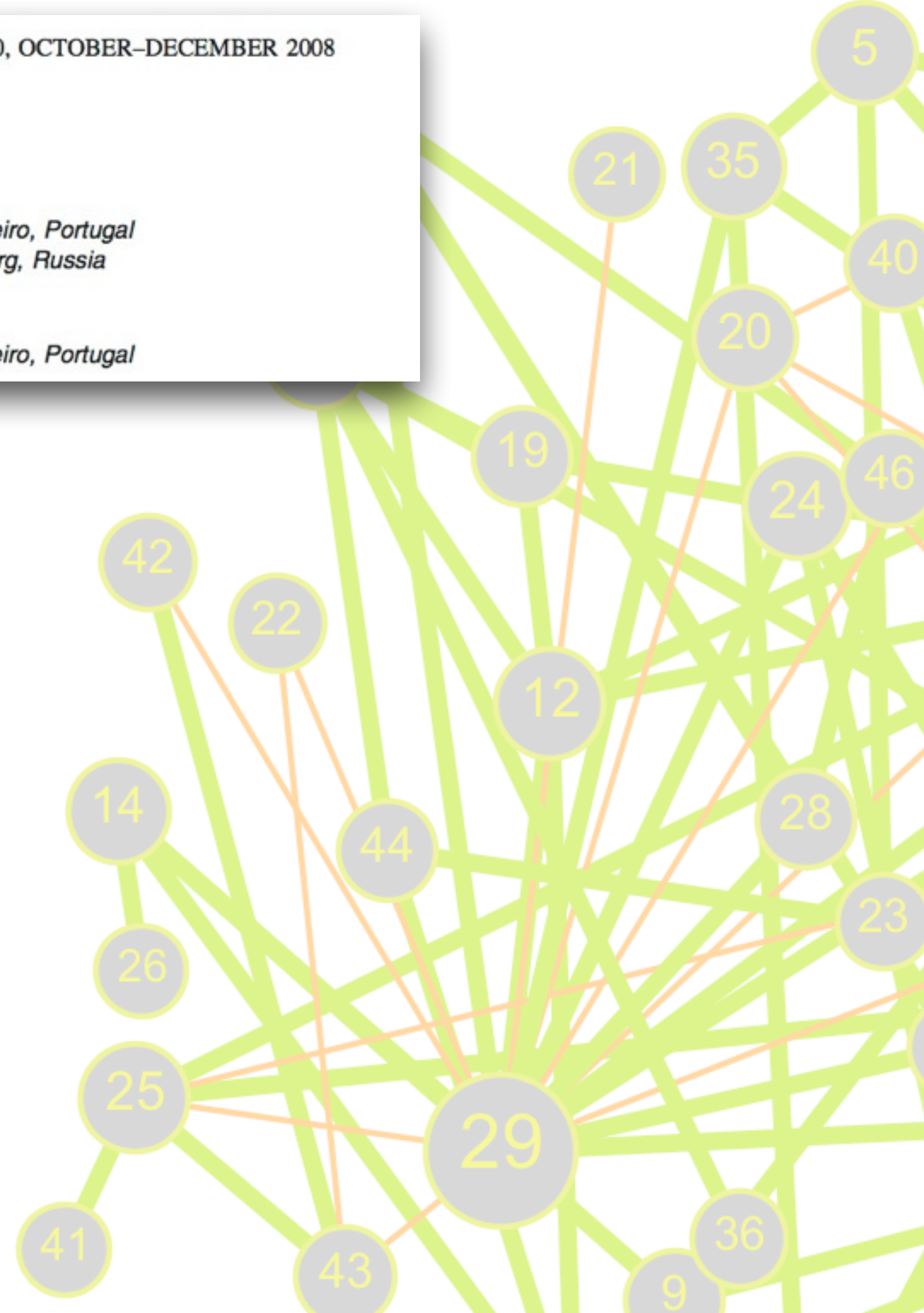
Critical phenomena in complex networks

S. N. Dorogovtsev[*] and A. V. Goltsev[†]

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal
and A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia

J. F. F. Mendes[‡]

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal



@gomezgardenes

REVIEWS OF MODERN PHYSICS, VOLUME 80, OCTOBER–DECEMBER 2008

Critical phenomena in complex networks

S. N. Dorogovtsev[*] and A. V. Goltsev[†]

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal and A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia

J. F. F. Mendes[‡]

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal

REVIEWS OF MODERN PHYSICS, VOLUME 81, APRIL–JUNE 2009

Statistical physics of social dynamics

Claudio Castellano[*]

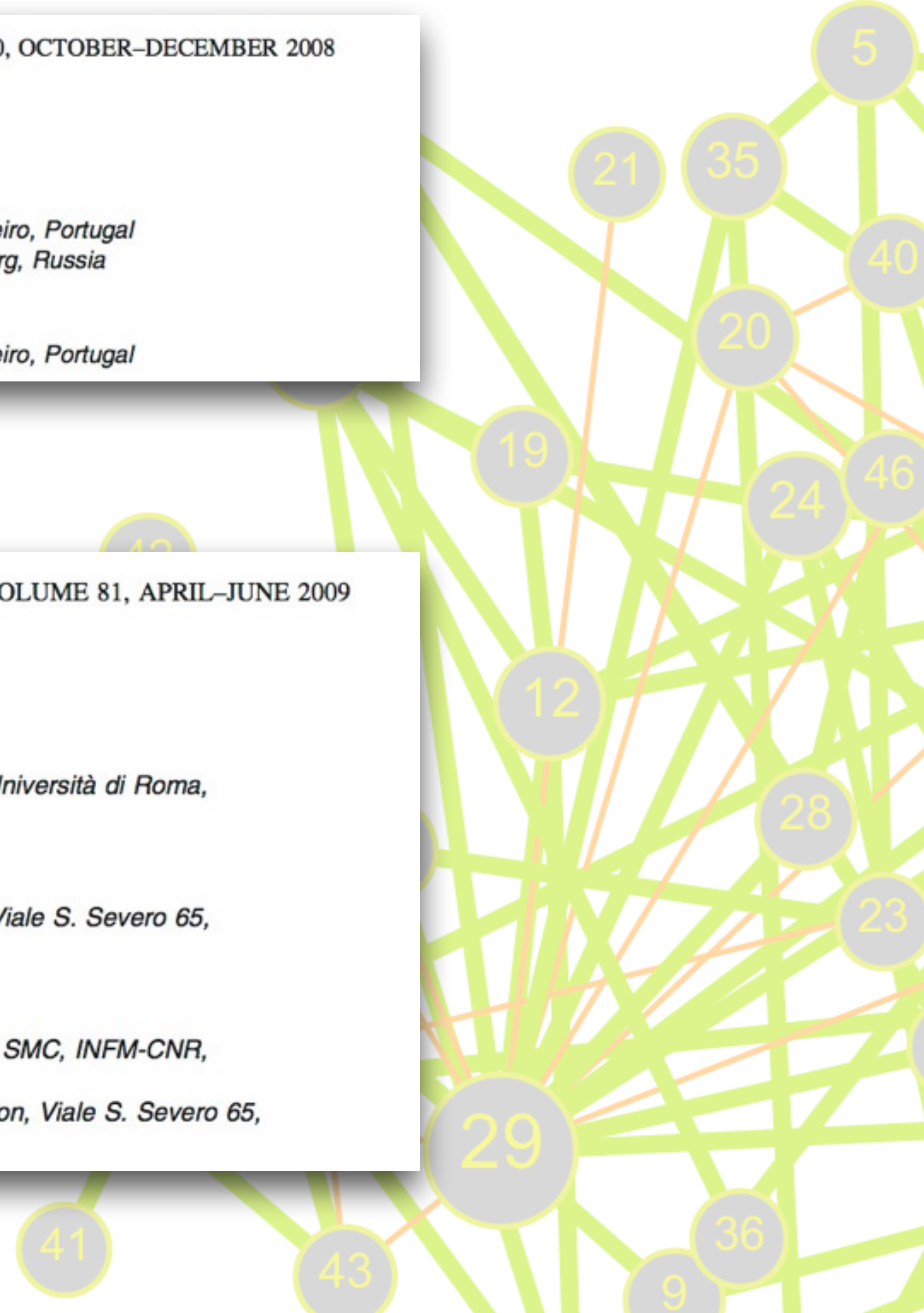
SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy

Santo Fortunato[†]

Complex Networks Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133 Torino, Italy

Vittorio Loreto[‡]

Dipartimento di Fisica, "Sapienza" Università di Roma and SMC, INFM-CNR, Piazzale A. Moro 2, 00185 Roma, Italy and Complex Networks Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133 Torino, Italy



@gomezgardenes

REVIEWS OF MODERN PHYSICS, VOLUME 80, OCTOBER–DECEMBER 2008

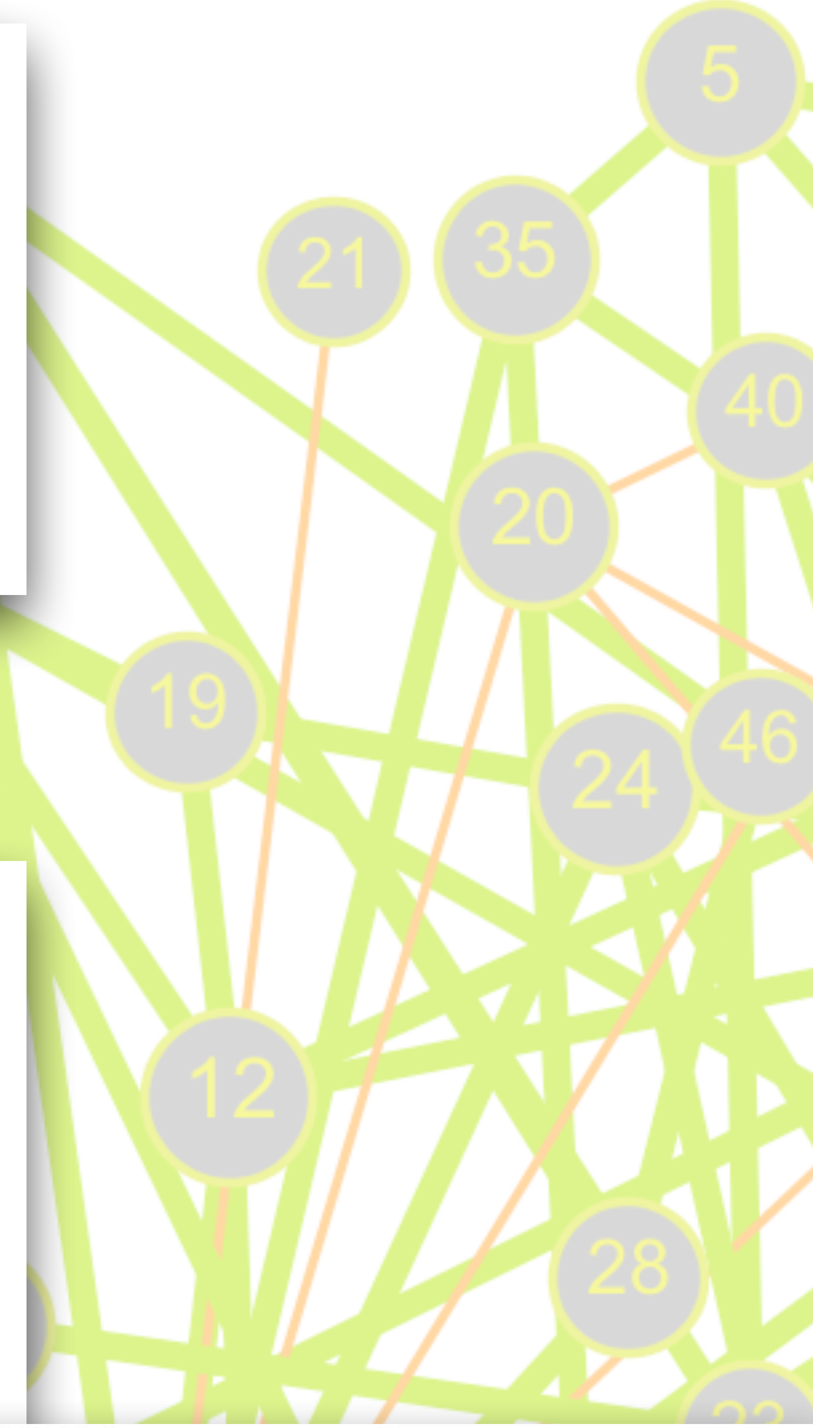
Critical phenomena in complex networks

S. N. Dorogovtsev* and A. V. Goltsev†

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal and A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia

J. F. F. Mendes‡

Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal



REVIEWS OF MODERN PHYSICS, VOLUME 81, APRIL–JUNE 2009

Statistical physics of social dynamics

Claudio Castellano*

SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy

Santo Fortunato†

Complex Networks Lagrange La
10133 Torino, Italy

Vittorio Loreto‡

Dipartimento di Fisica, "Sapienza
Piazzale A. Moro 2, 00185 Roma
and Complex Networks Lagrang
10133 Torino, Italy

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

PHYSICS REPORTS

Physics Reports 446 (2007) 97–216

www.elsevier.com/locate/physrep

Evolutionary games on graphs

György Szabó[a,*], Gábor Fáth[b]

@gomezgardenes

REVIEWS OF MODERN PHYSICS, VOLUME 80, OCTOBER–DECEMBER 2008

Critical phenomena in complex networks

S. N. Dorogovtsev* and A. V. Goltsev†

Departamento de Física, Universidade de Aveiro, 3
and A. F. Ioffe Physico-Technical Institute, 194021

J. F. F. Mendes‡

Departamento de Física, Universidade de Aveiro, 3

Physics Reports 469 (2008) 93–153

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

ELSEVIER

Synchronization in complex networks

Alex Arenas[a,b], Albert Díaz-Guilera[c,b], Jurgen Kurths[d,e], Yamir Moreno[b,f,*], Changsong Z

REVIEWS OF MODERN PHYSICS, VOLUME 81, APRIL–JUNE 2009

Statistical physics of social dynamics

Claudio Castellano*

SMC, INFM-CNR and Dipartimento di Fisica, "Sapienza" Università di Roma,
Piazzale A. Moro 2, 00185 Roma, Italy

Santo Fortunato†

Complex Networks Lagrange La
10133 Torino, Italy

Vittorio Loreto‡

Dipartimento di Fisica, "Sapienza
Piazzale A. Moro 2, 00185 Roma
and Complex Networks Lagrang
10133 Torino, Italy

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

PHYSICS REPORTS

Physics Reports 446 (2007) 97–216

www.elsevier.com/locate/physrep

Evolutionary games on graphs

György Szabó[a,*], Gábor Fáth[b]

@gomezgardenes

REVIEWS OF MODERN PHYSICS, VOLUME 80, OCTOBER–DECEMBER 2008

Critical phenomena in complex networks

S. N. Dorogovtsev* and A. V. Goltsev†

Departamento de Física, Universidade de Aveiro, 3
and A. F. Ioffe Physico-Technical Institute, 194021

J. F. F. Mendes‡

Departamento de Física, Universidade de Aveiro, 3

Physics Reports 469 (2008) 93–153

Contents lists available at ScienceDirect

Physics Reports

journal homepage: www.elsevier.com/locate/physrep

REVIEWS OF MODERN PHYSICS, VOLUME 87, JULY–SEPTEMBER 2015

Epidemic processes in complex networks

Romualdo Pastor-Satorras*

Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

Claudio Castellano

Istituto dei Sistemi Complessi (ISC-CNR), via dei Taurini 19, I-00185 Roma, Italy and Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy

Piet Van Mieghem

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Alessandro Vespignani

Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston Massachusetts 02115, USA and Institute for Scientific Interchange Foundation, Turin 10133, Italy

Statistical ph

Claudio C

SMC, INFM
Piazzale A.

Santo For

Complex N
10133 Tori

Vittorio Lo

Dipartimen
Piazzale A. Moro 2, 00185 Roma
and Complex Networks Lagrang
10133 Torino, Italy

mir Moreno[b,f,*], Changsong Z

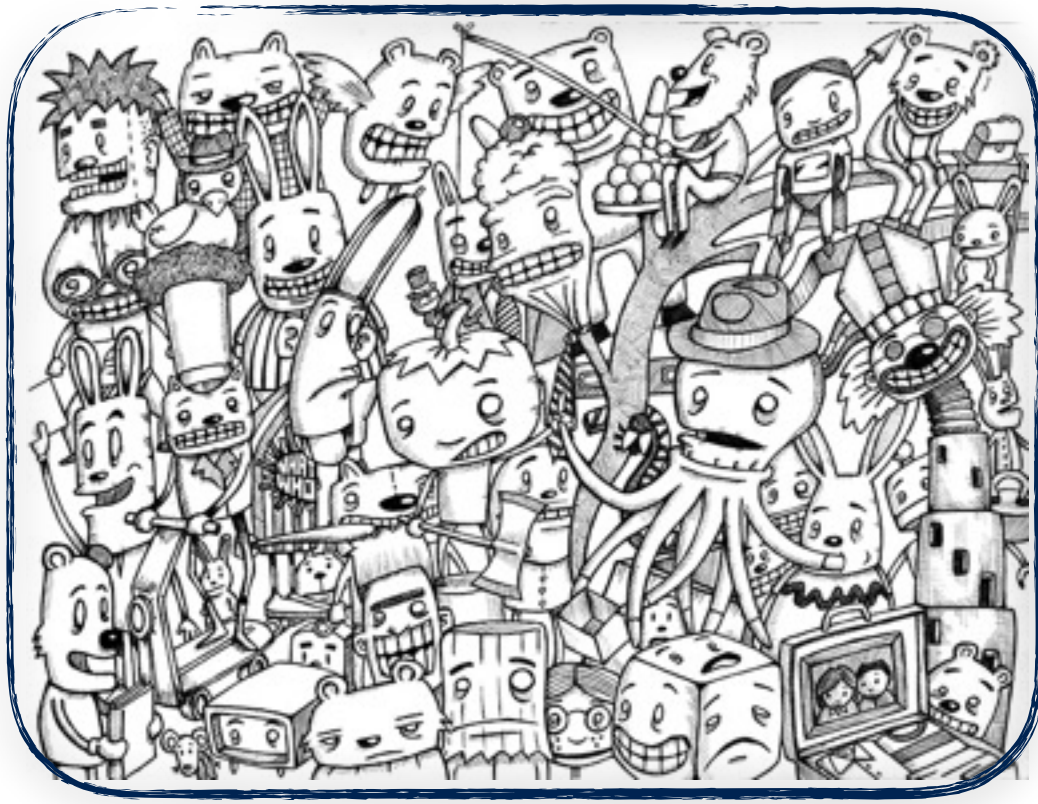
PHYSICS REPORTS

www.elsevier.com/locate/physrep

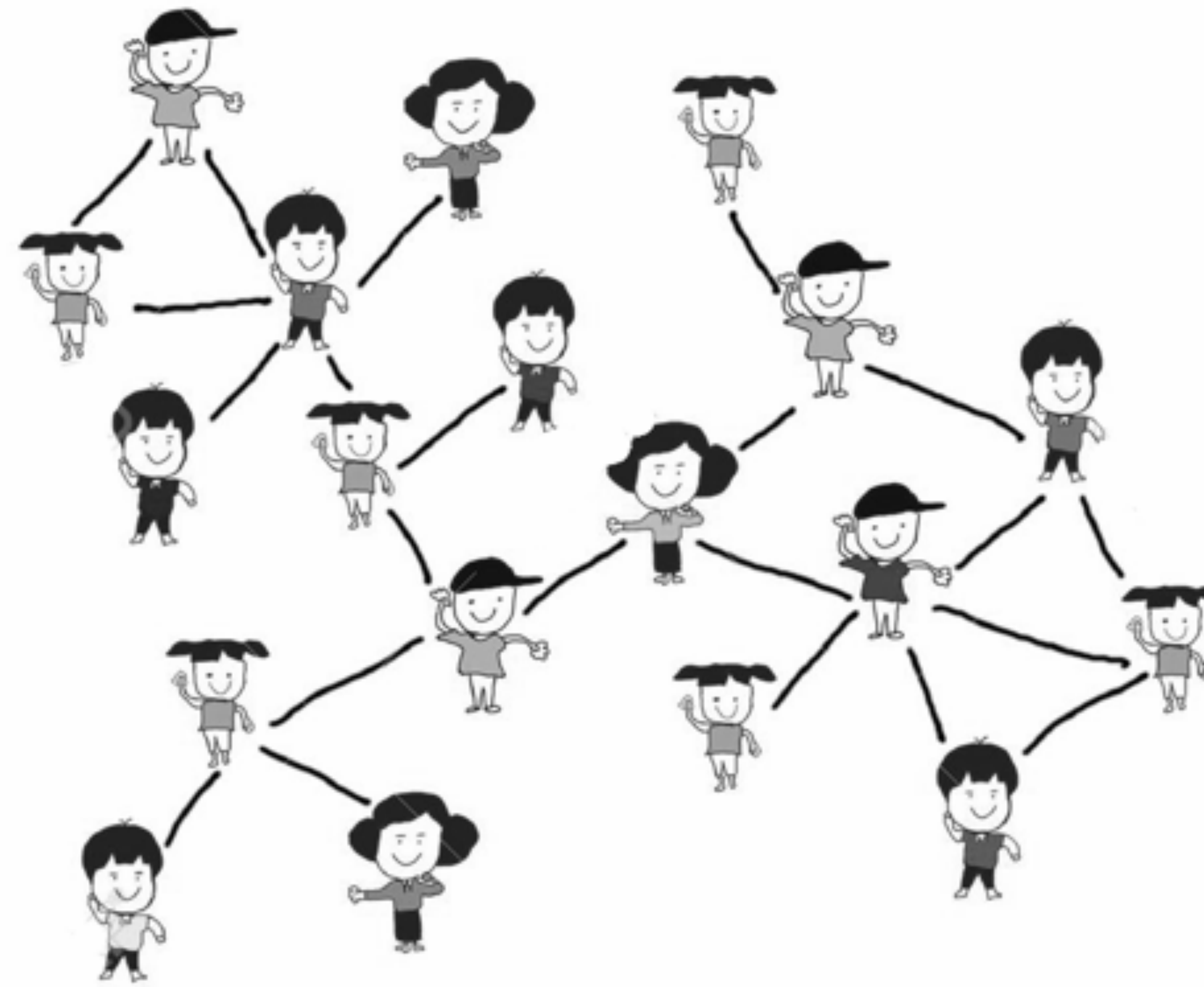
Evolutionary games on graphs

György Szabó[a,*], Gábor Fáth[b]

@gomezgardenes

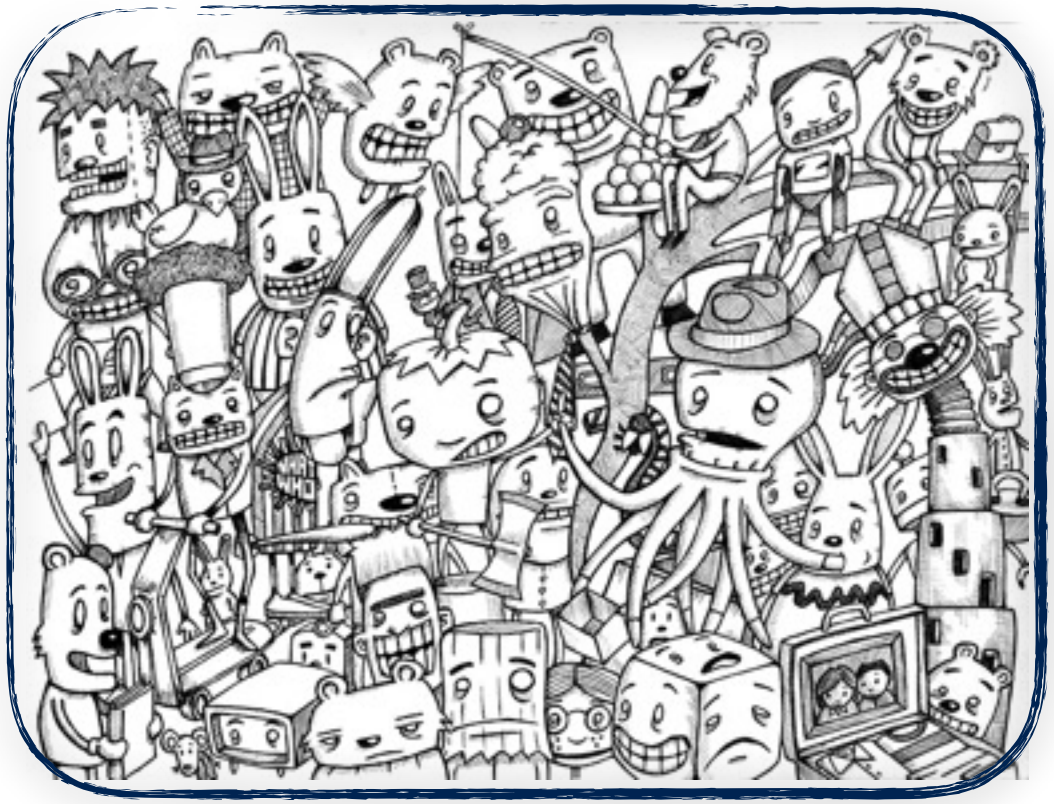


+ Realism
+ Complexity



+ Realism
+ Complexity

@gomezgardenes



Well mixed / Mean field

Compartmental models

Aimed at capturing the global (population-level) dynamics from the microscopic contagion processes

Each individual can be in one of *n* states at time *t*

S - Susceptible (Healty)

I - Infected (and infectious)

R - Recovered (immune/dead)

(From Petter Holme's blog)

Compartmental models

The transitions (e.g. S $\rightarrow$I) are mediated by some rates:
λ and μ

$\xrightarrow{\lambda}$

$\xrightarrow{\mu}$

S - Susceptible (Healty) **I** - Infected (and infectious) **R** - Recovered (immune/dead)

@gomezgardenes

(From Petter Holme's blog)

Compartmental models

The transitions (e.g. S $\longrightarrow$I) are mediated by some rates:
λ and μ

The final impact of an SIR epidemic is given by the fraction of affected (Recovered) individuals

$\xrightarrow{\lambda}$

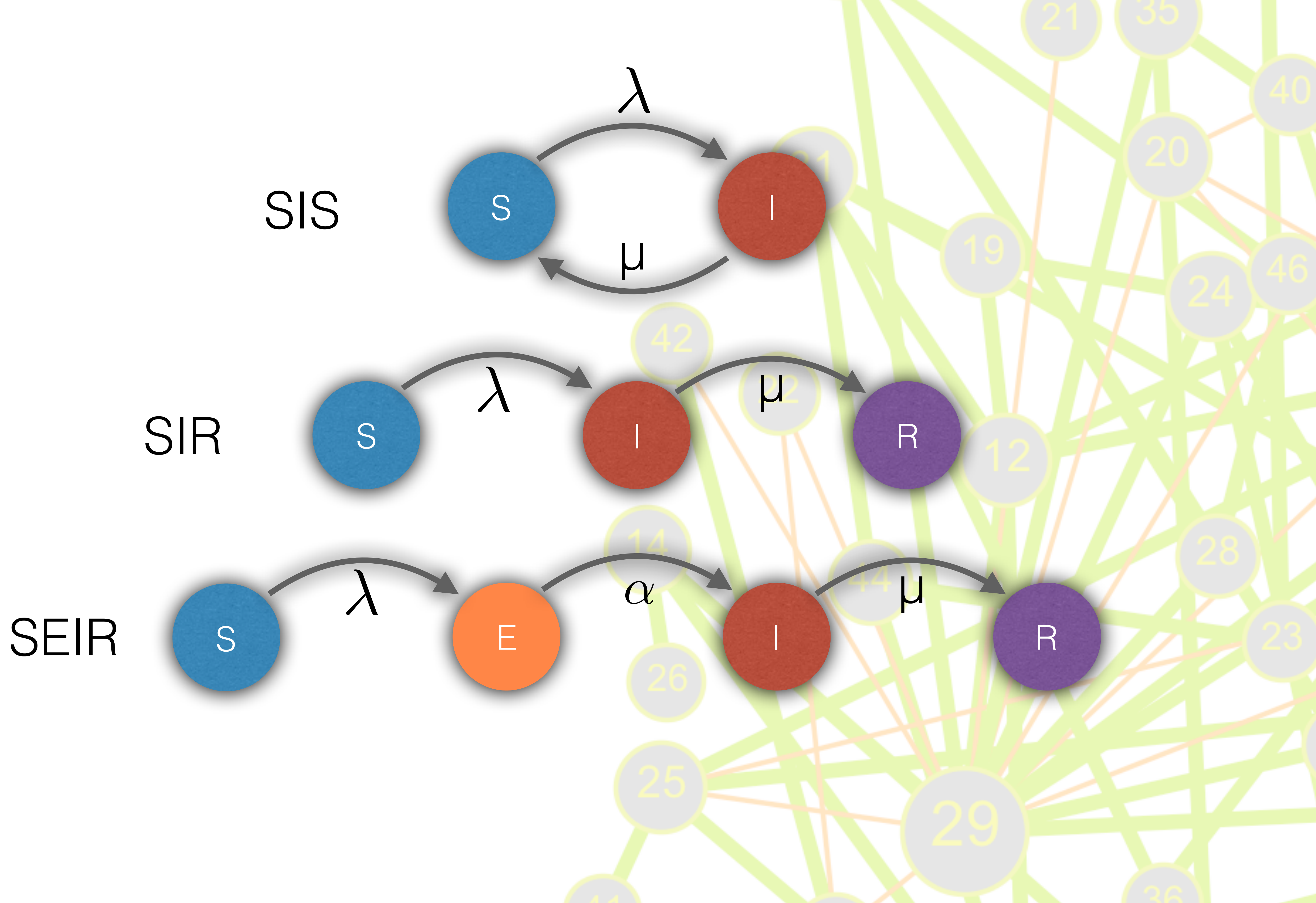
$\xrightarrow{\mu}$

S - Susceptible (Healty) **I** - Infected (and infectious) **R** - Recovered (immune/dead)

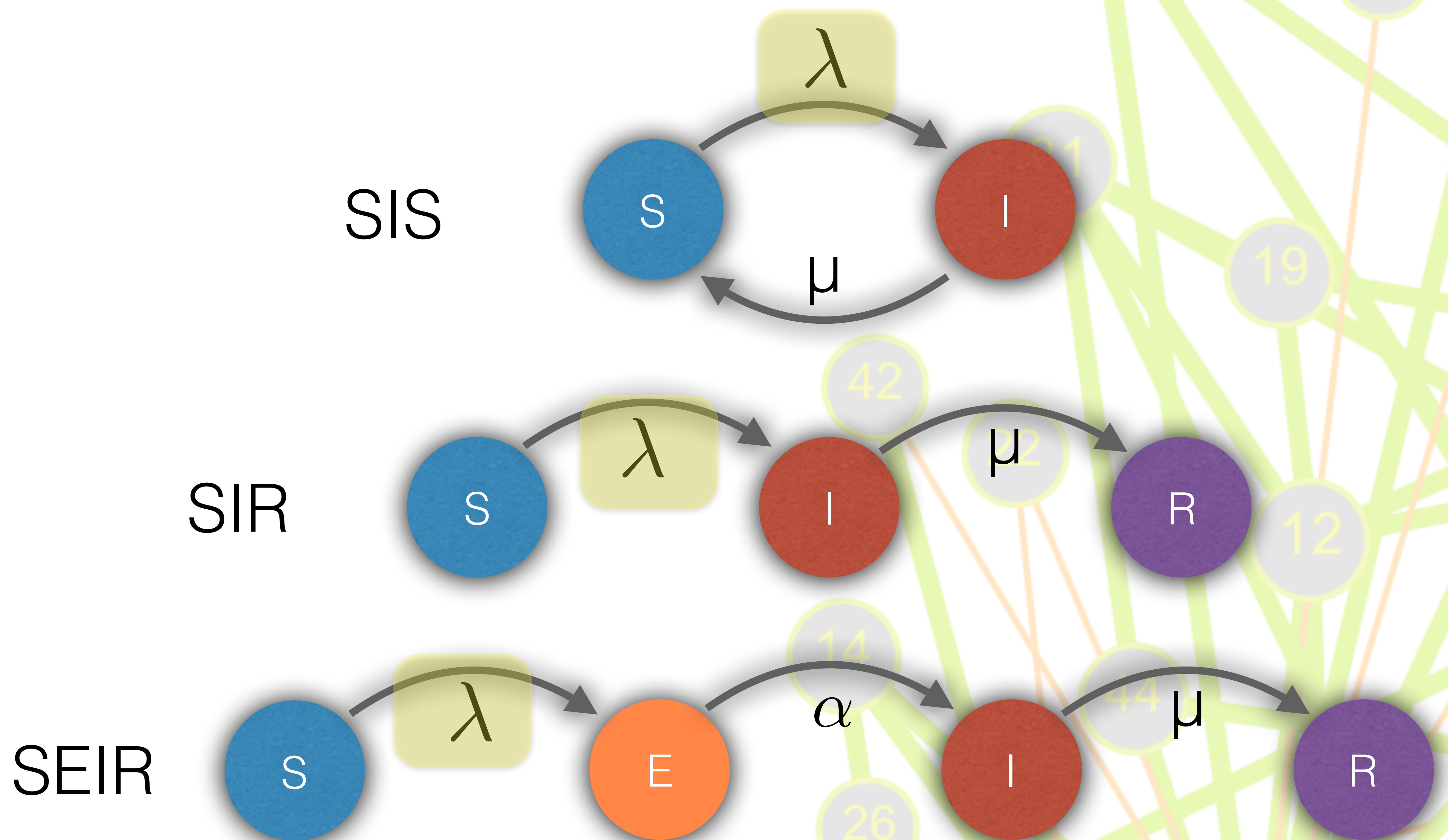
@gomezgardenes

(From Petter Holme's blog)

Some examples



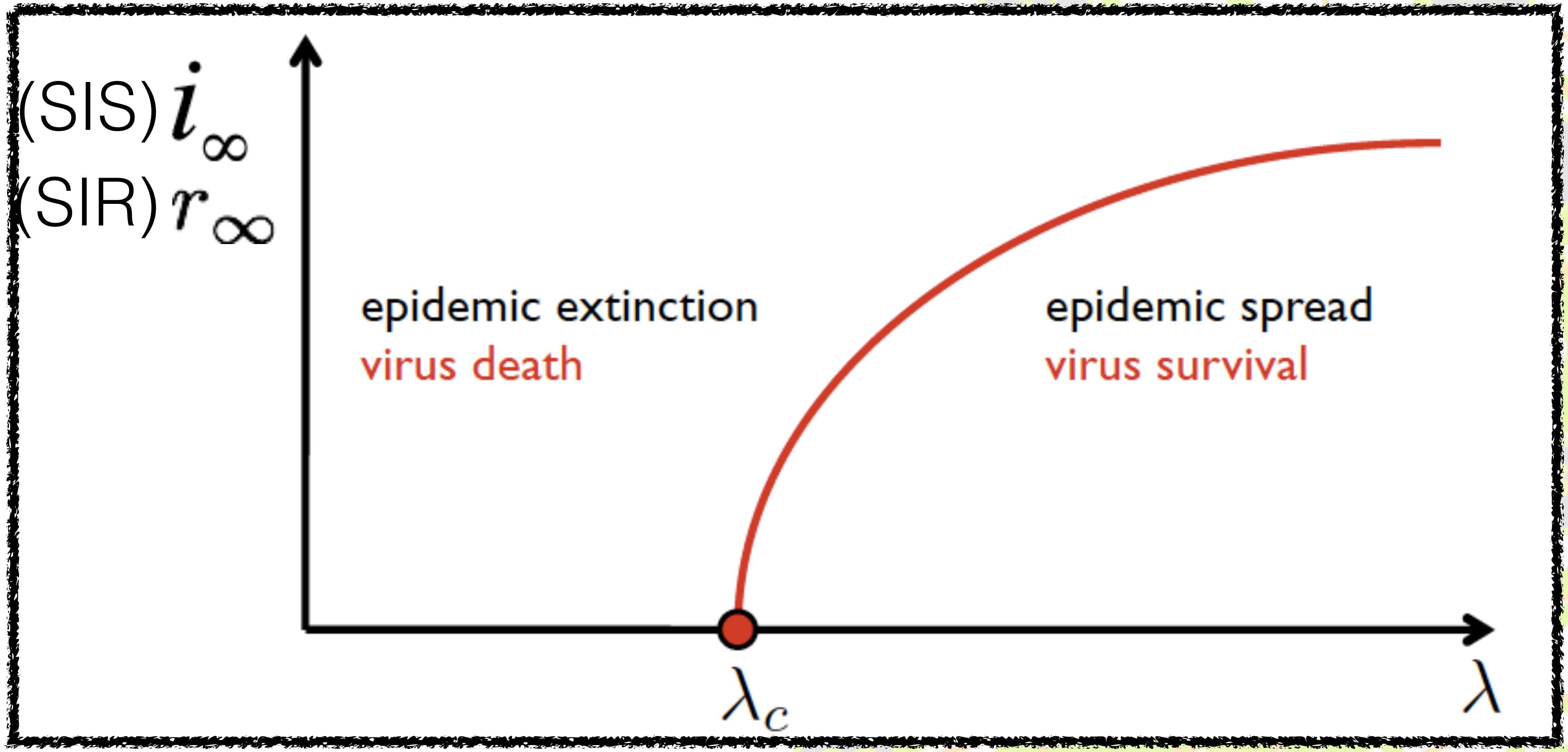
Some examples



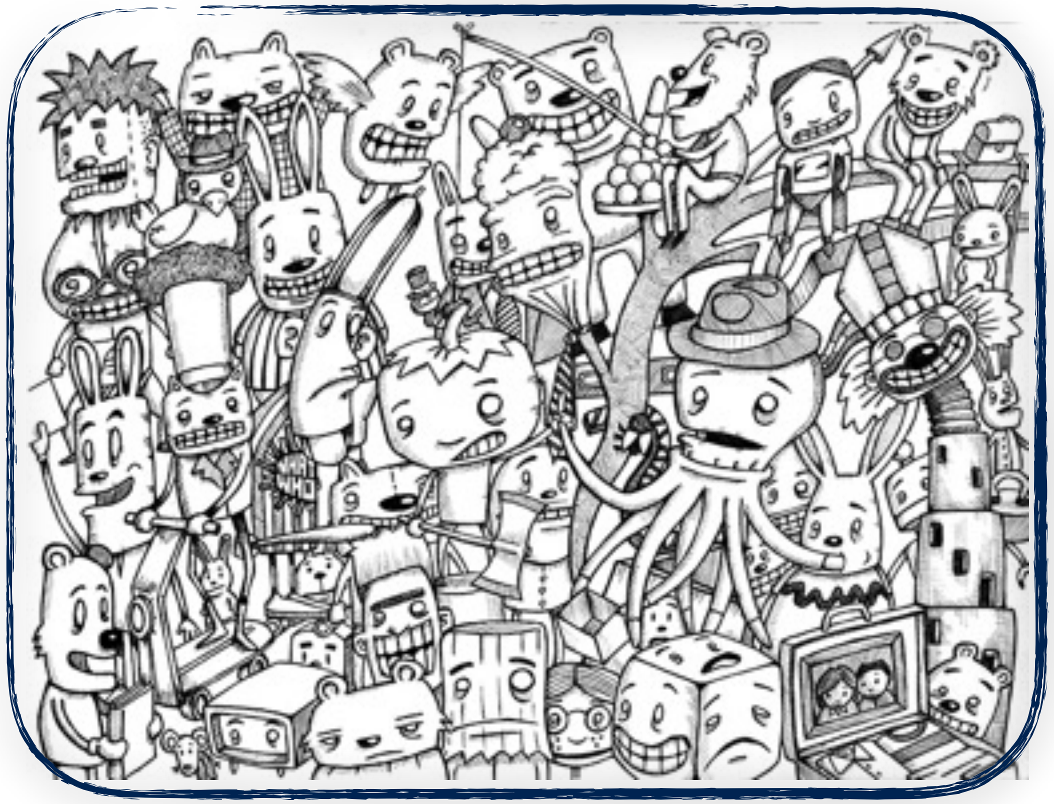
QUESTION: What is the minimum value of λ for the epidemic outbreak to take place?

@gomezgardenes

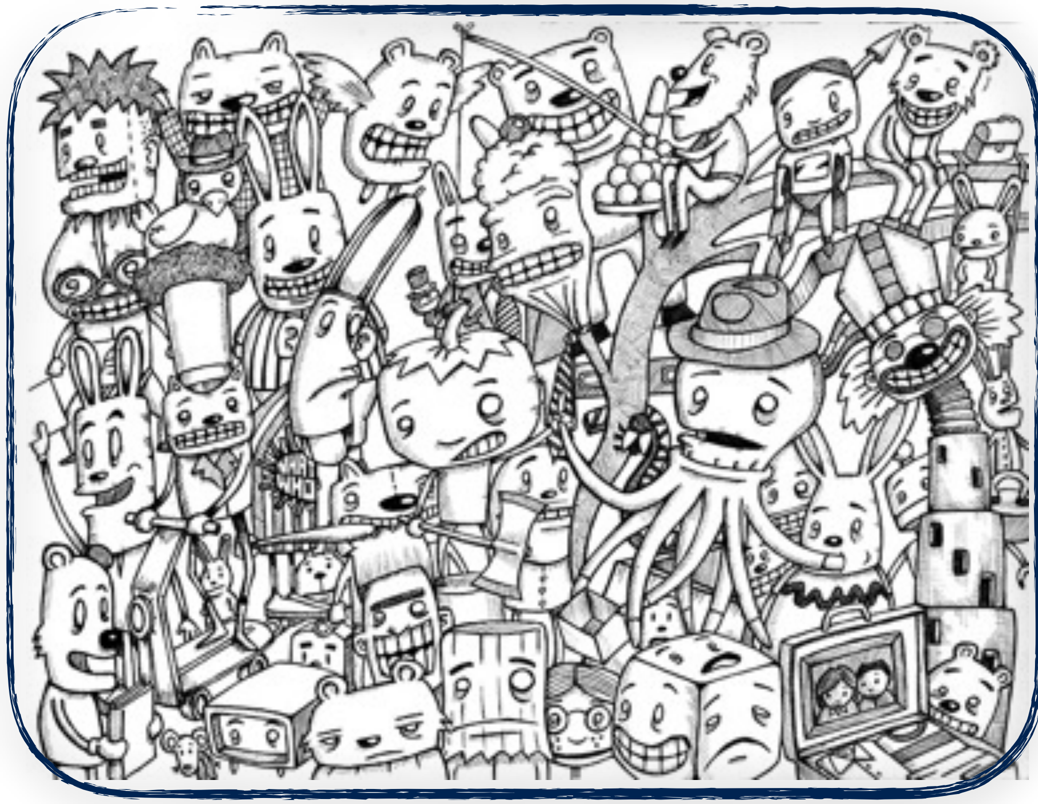
Epidemic Threshold



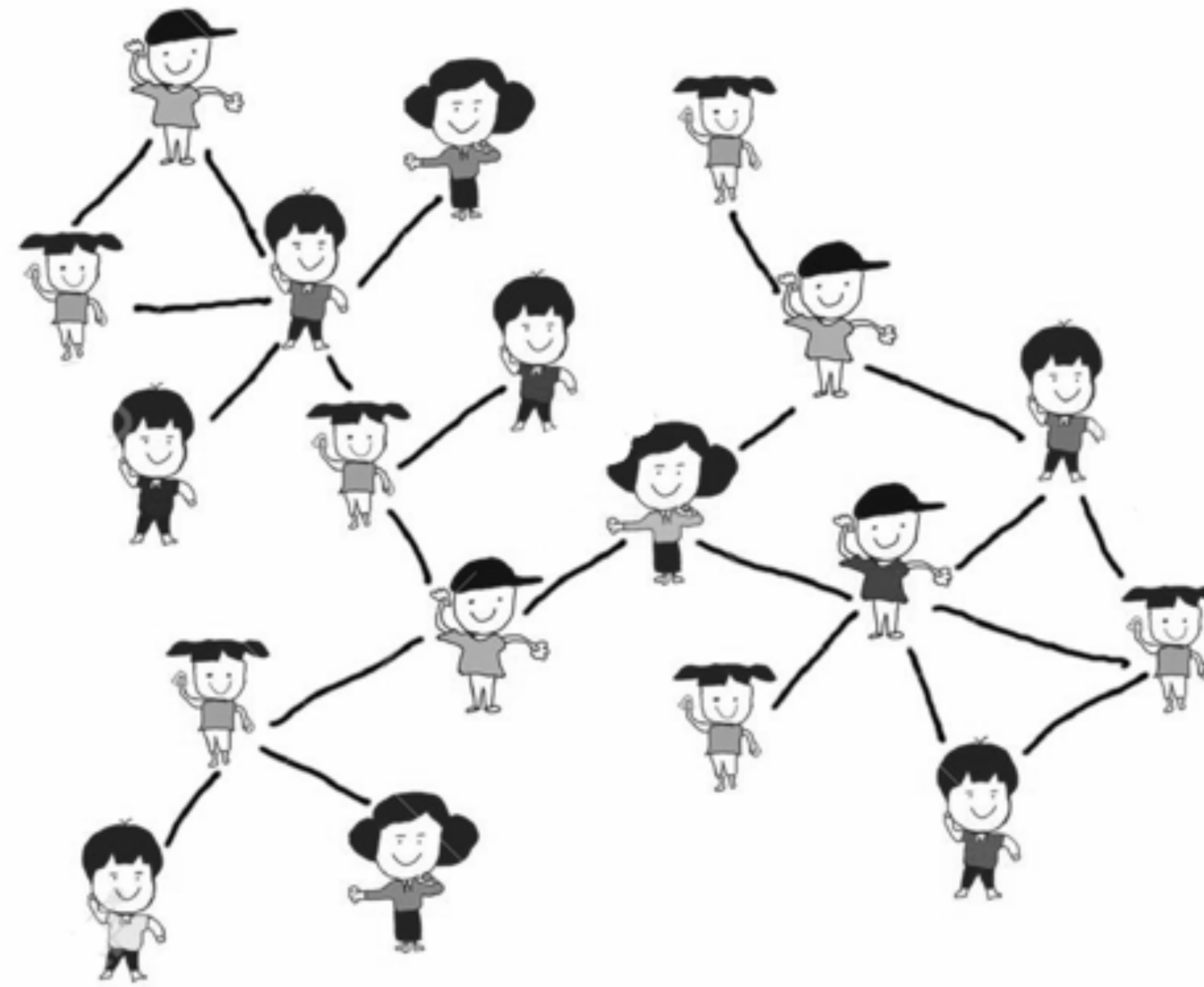
@gomezgardenes



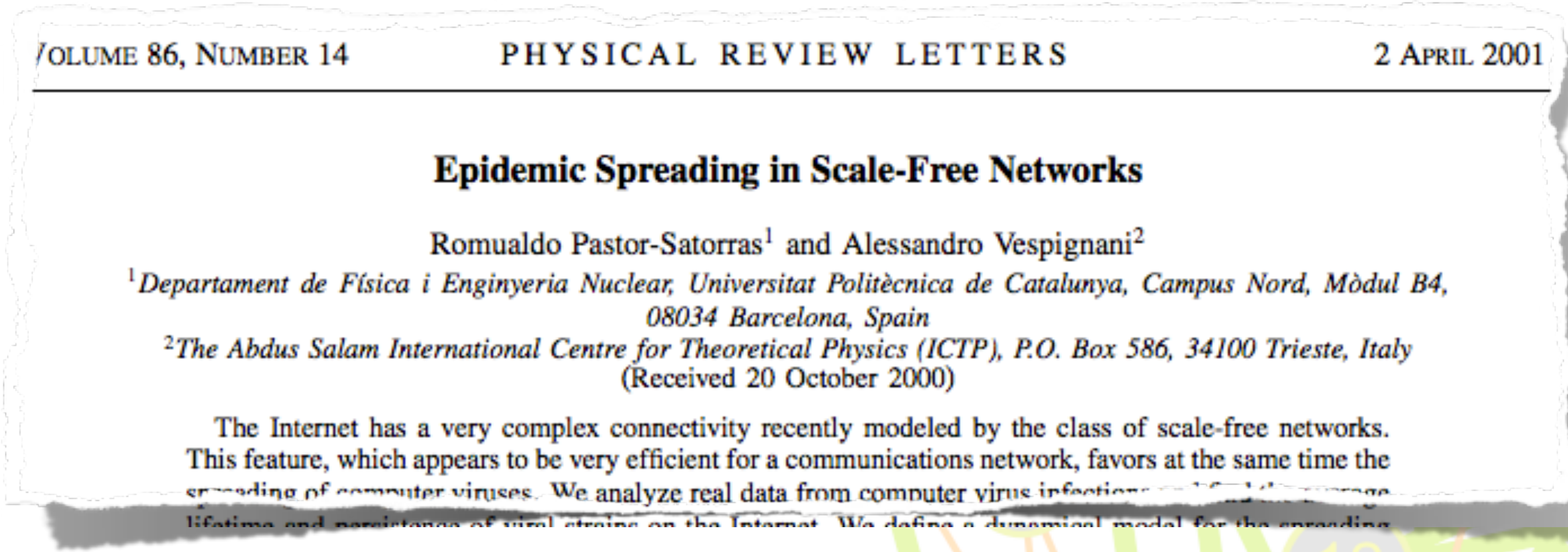
Well mixed / Mean field



+ Realism
+ Complexity



Epidemics & Networks



VOLUME 86, NUMBER 14 PHYSICAL REVIEW LETTERS 2 APRIL 2001

Epidemic Spreading in Scale-Free Networks

Romualdo Pastor-Satorras[1] and Alessandro Vespignani[2]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*
[2]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*

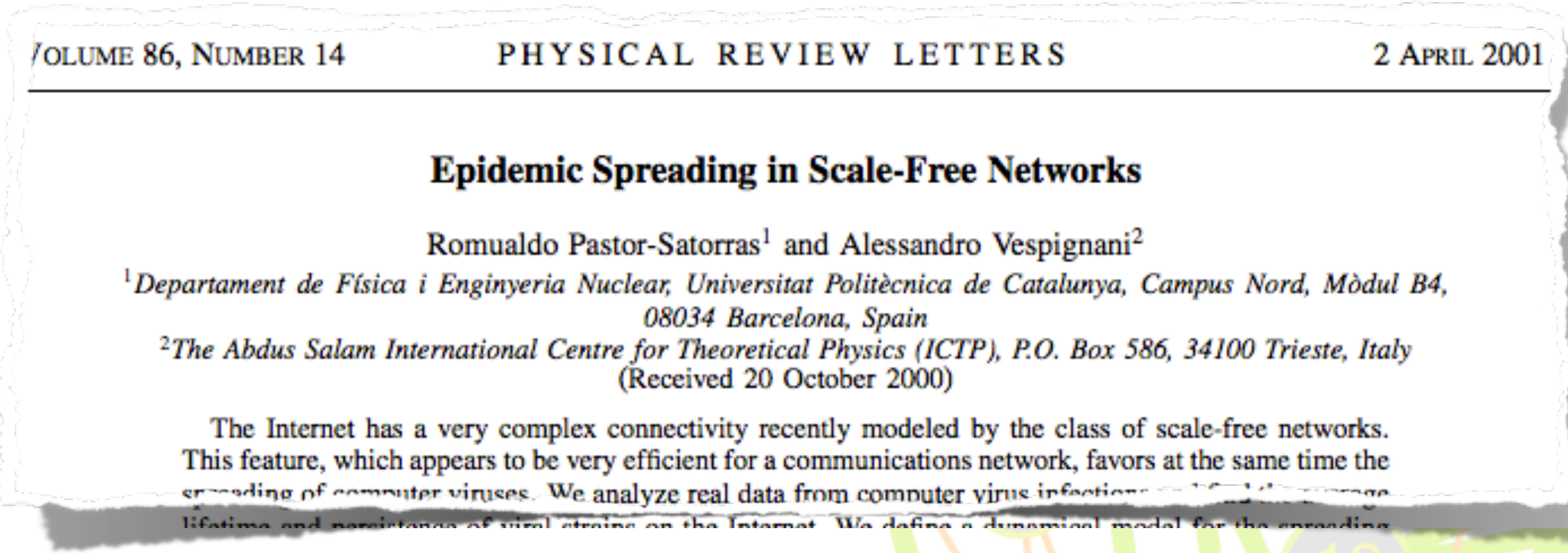
(Received 20 October 2000)

The Internet has a very complex connectivity recently modeled by the class of scale-free networks. This feature, which appears to be very efficient for a communications network, favors at the same time the ... of computer viruses. We analyze real data from computer virus

Google scholar More than 3900 citations

@gomezgardenes

Epidemics & Networks



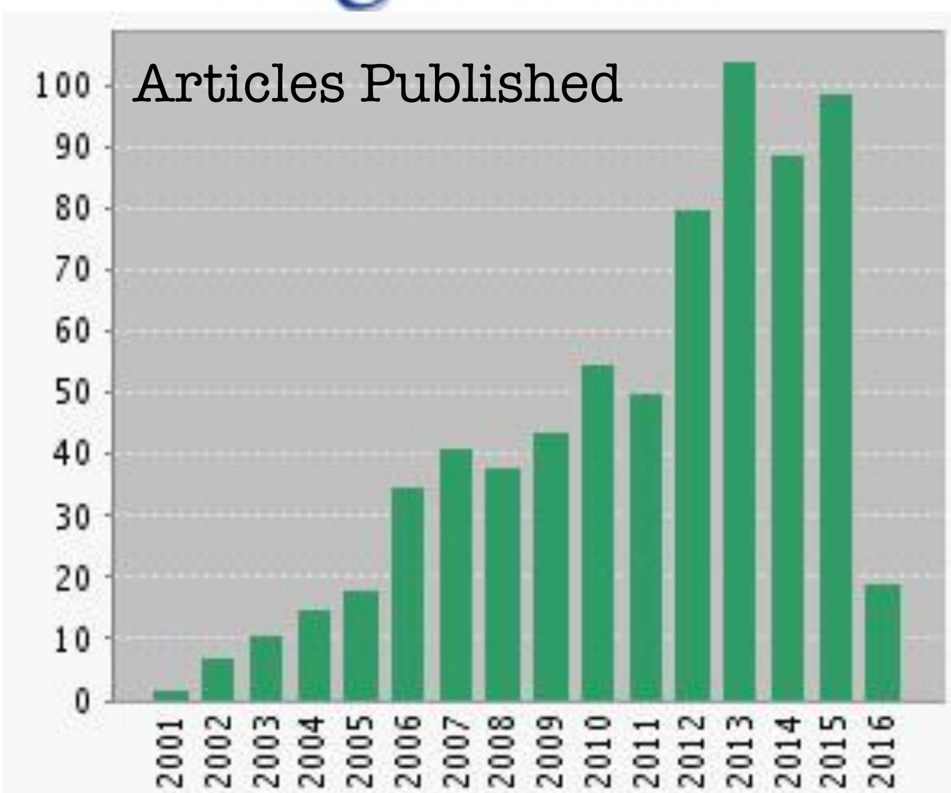
/OLUME 86, NUMBER 14 PHYSICAL REVIEW LETTERS 2 APRIL 2001

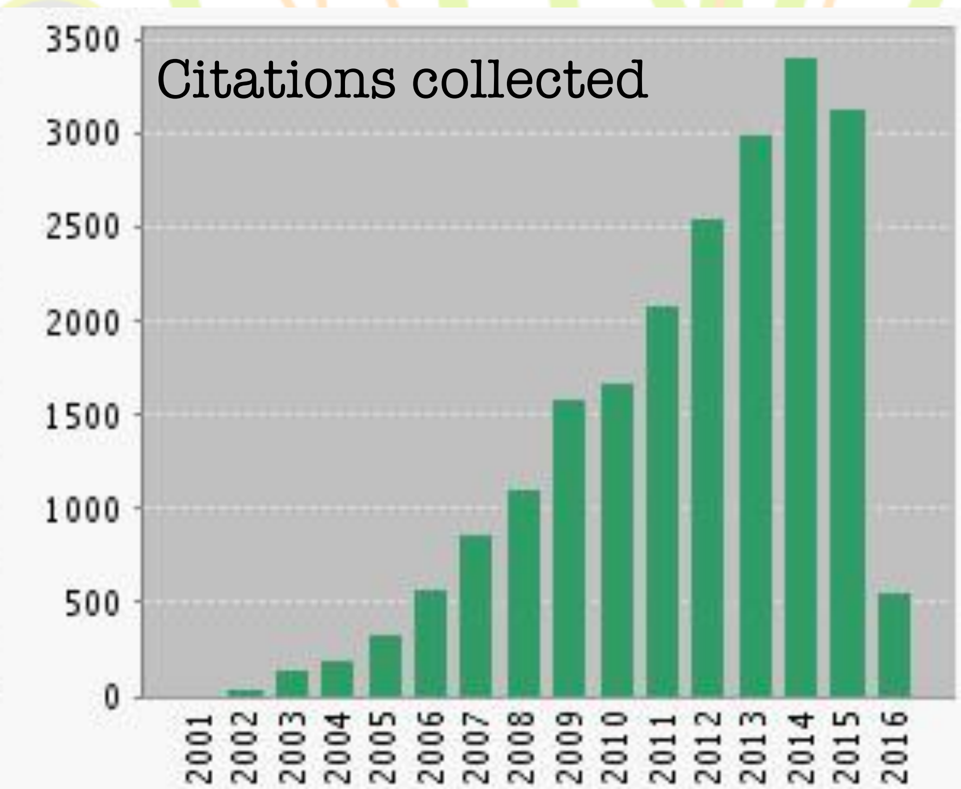
Epidemic Spreading in Scale-Free Networks

Romualdo Pastor-Satorras[1] and Alessandro Vespignani[2]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*
[2]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*
(Received 20 October 2000)

The Internet has a very complex connectivity recently modeled by the class of scale-free networks. This feature, which appears to be very efficient for a communications network, favors at the same time the

Google scholar More than 3900 citations

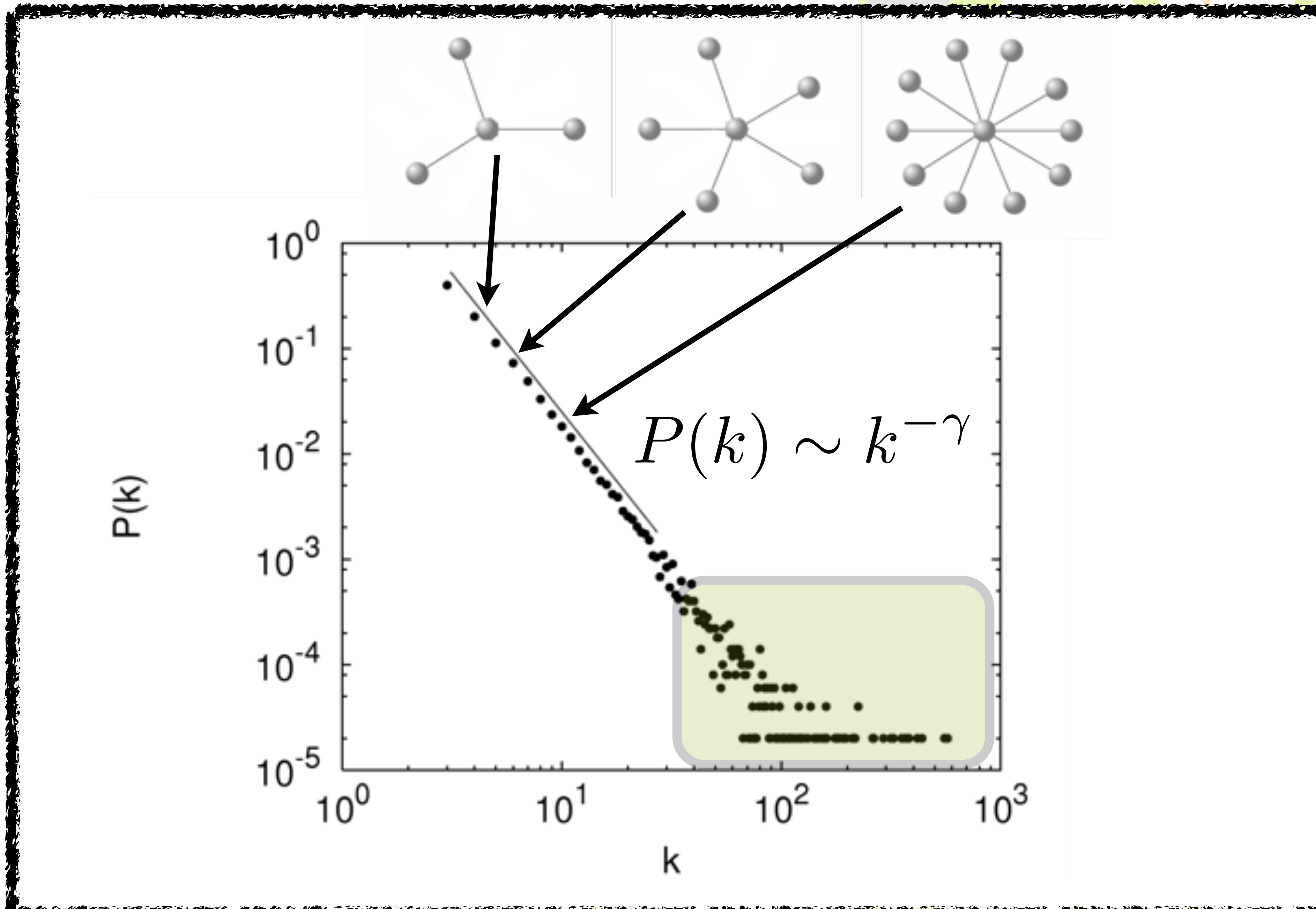




@gomezgardenes

Epidemics & Networks

Scale-free phenomenon



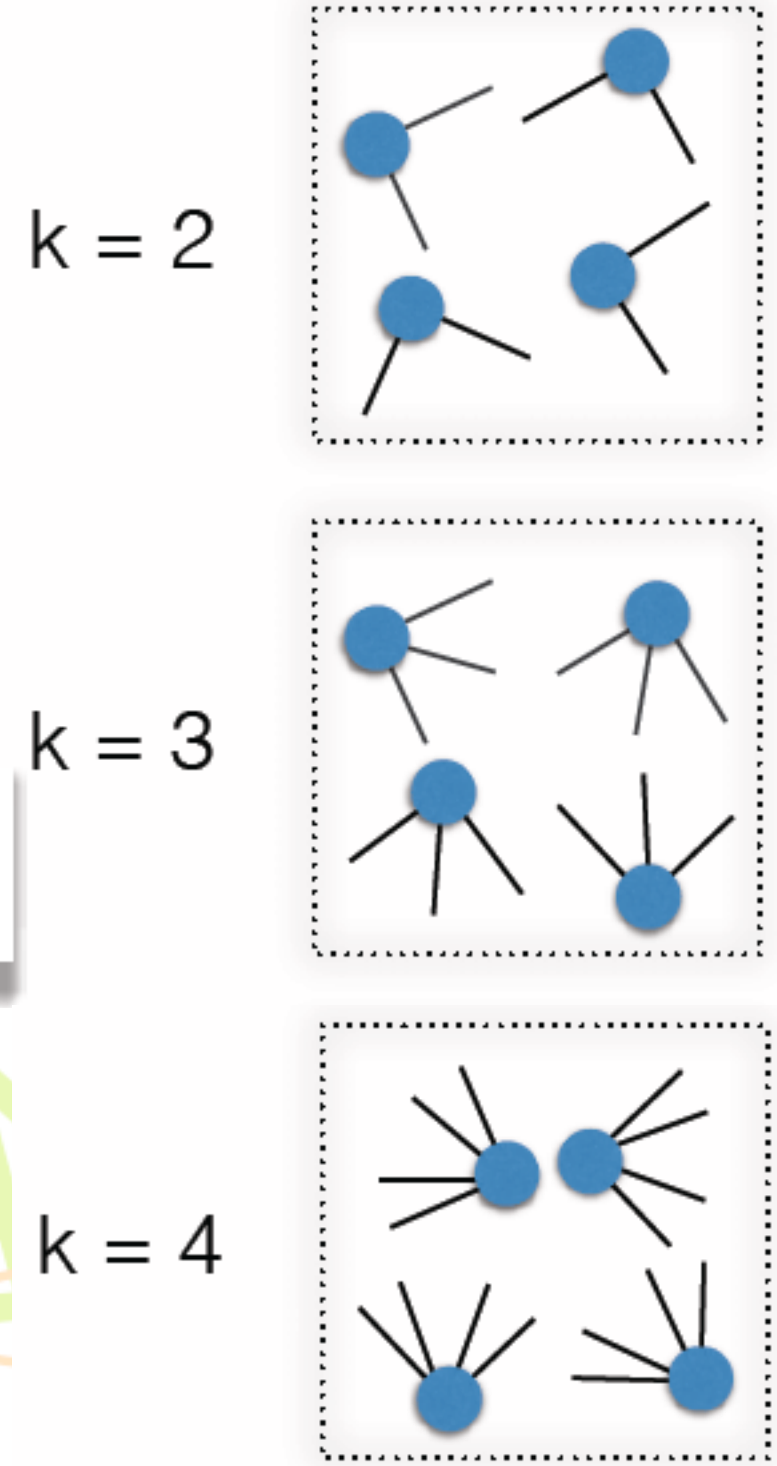
@gomezgardenes

Epidemics & Networks

In heterogeneous networks
the approximation $k \sim \langle k \rangle$
doesn't hold

Solution:
Degree Block approximation

~~All the nodes are statistically equivalent~~
All the nodes with the same degree are statistically equivalent



@gomezgardenes

Epidemics & Networks

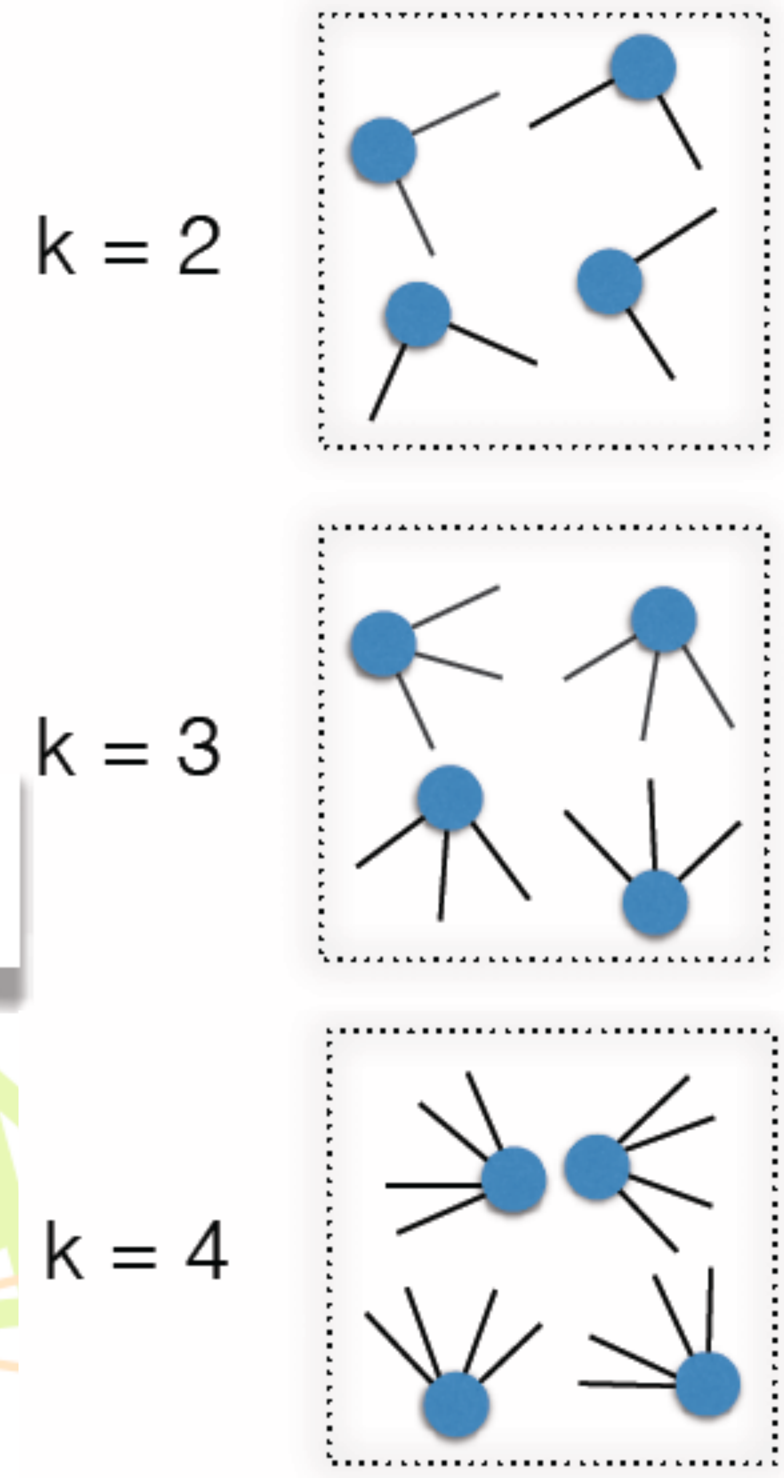
In heterogeneous networks the approximation $k \sim \langle k \rangle$ doesn't hold

Solution:
Degree Block approximation

~~All the nodes are statistically equivalent~~
All the nodes with the same degree are statistically equivalent

$$i_k = \frac{I_k}{N_k} \qquad s_k = \frac{S_k}{N_k}$$

$$i = \sum_k P(k) i_k \qquad s = \sum_k P(k) s_k$$



@gomezgardenes

Epidemics & Networks

$$\boxed{\lambda_c \sim \frac{\langle k \rangle}{\langle k^2 \rangle}}$$

- in **Well-mixed** populations:

$$\langle k^2 \rangle = \langle k \rangle^2 \longrightarrow \lambda_c \sim \frac{1}{\langle k \rangle}$$

- in **Scale-Free** networks $P(k) \sim k^{-\gamma}$

if $2 < \gamma < 3$ then $\langle k^2 \rangle \to \infty$

$$\boxed{\lambda_c \to 0}$$

@gomezgardenes

Epidemics & Networks

$$\boxed{\lambda_c \sim \frac{\langle k\rangle}{\langle k^2\rangle}}$$

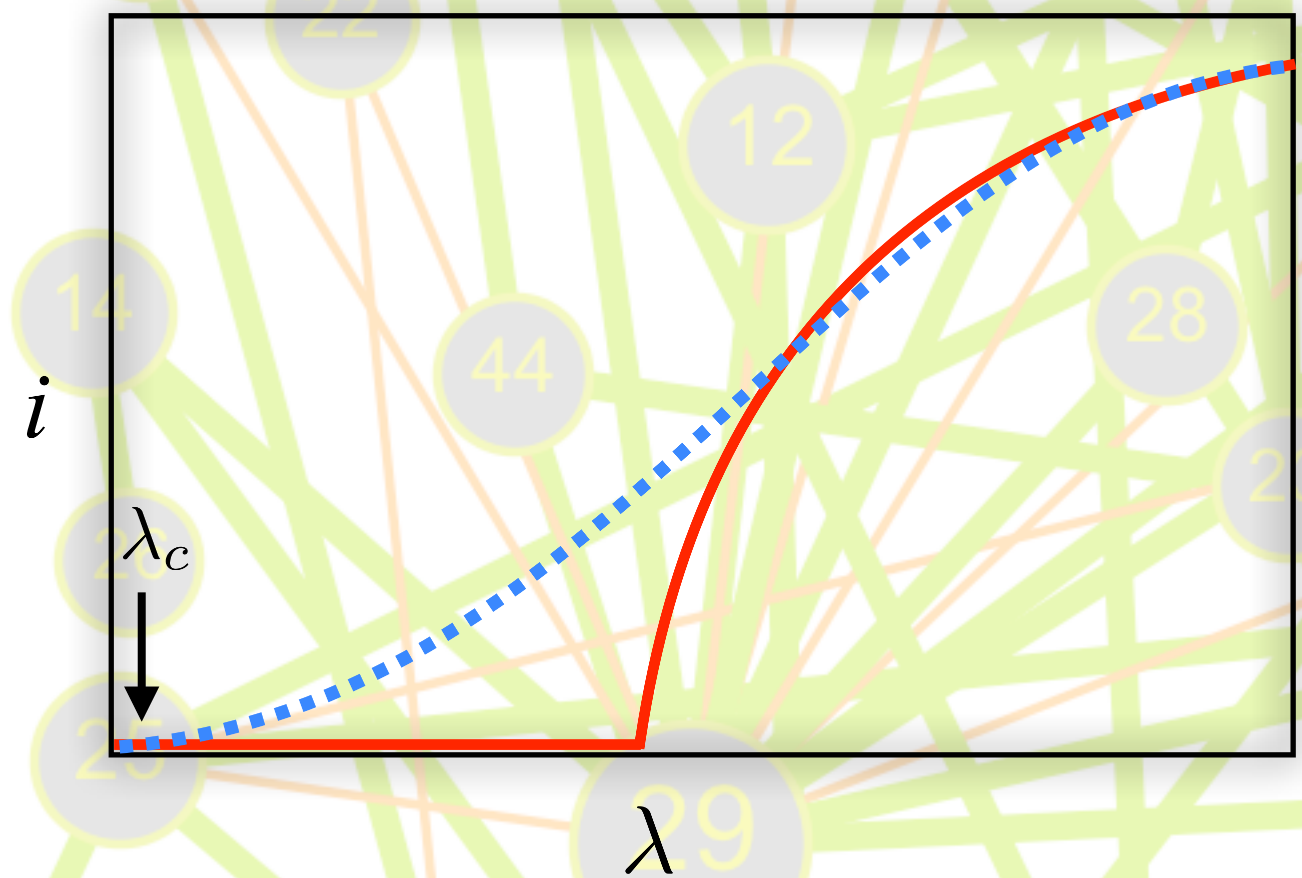
- in **Well-mixed** populations:

$$\langle k^2\rangle = \langle k\rangle^2 \longrightarrow \lambda_c \sim \frac{1}{\langle k\rangle}$$

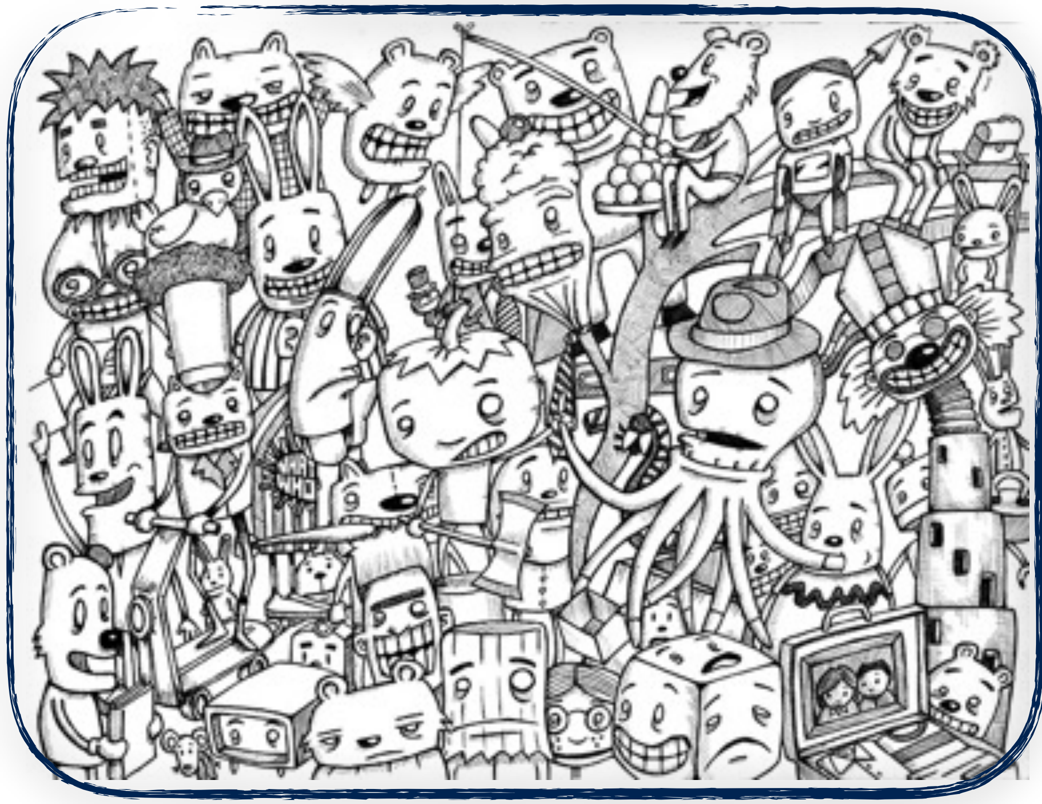
- in **Scale-Free** networks $P(k) \sim k^{-\gamma}$

  if $2 < \gamma < 3$ then $\langle k^2\rangle \to \infty$

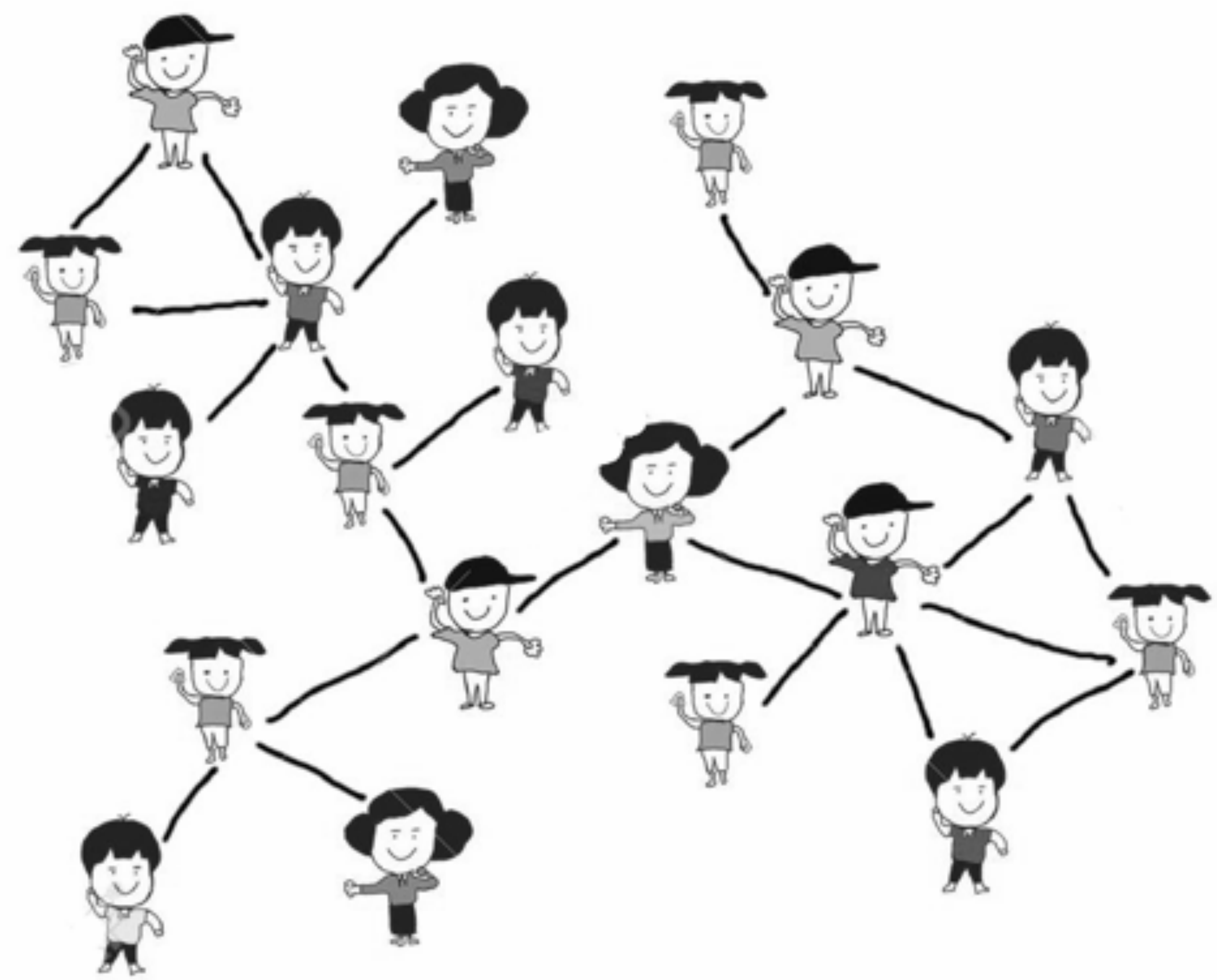
$$\boxed{\lambda_c \to 0}$$



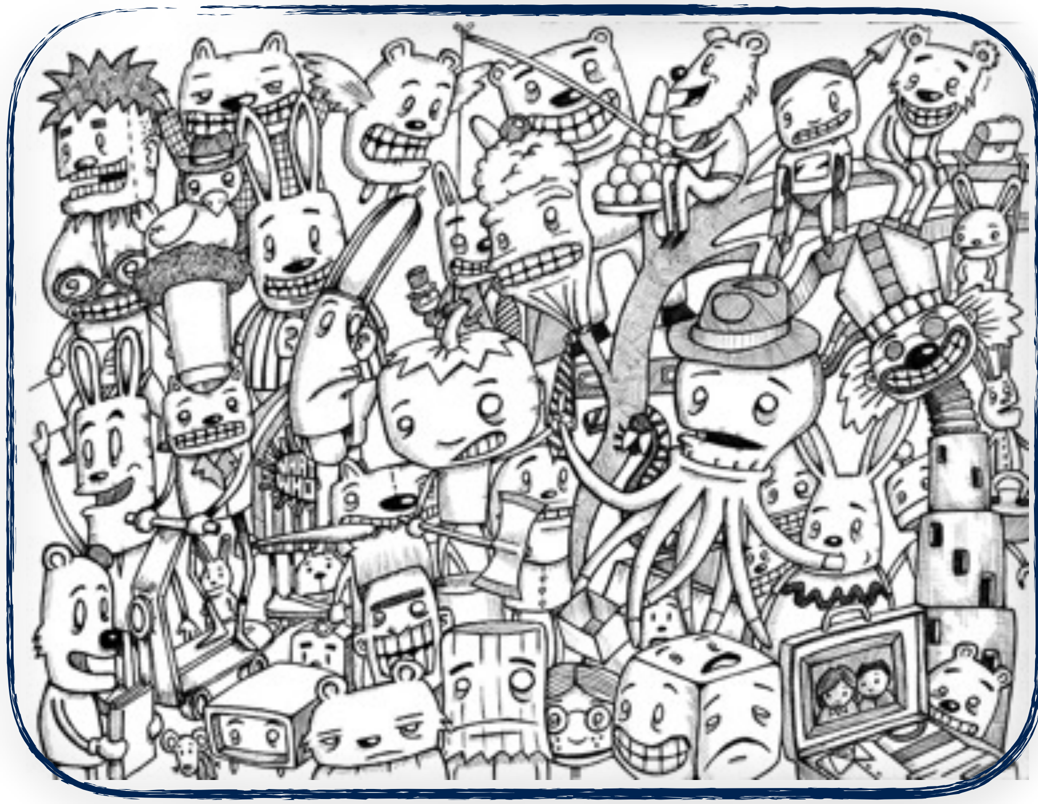
@gomezgardenes



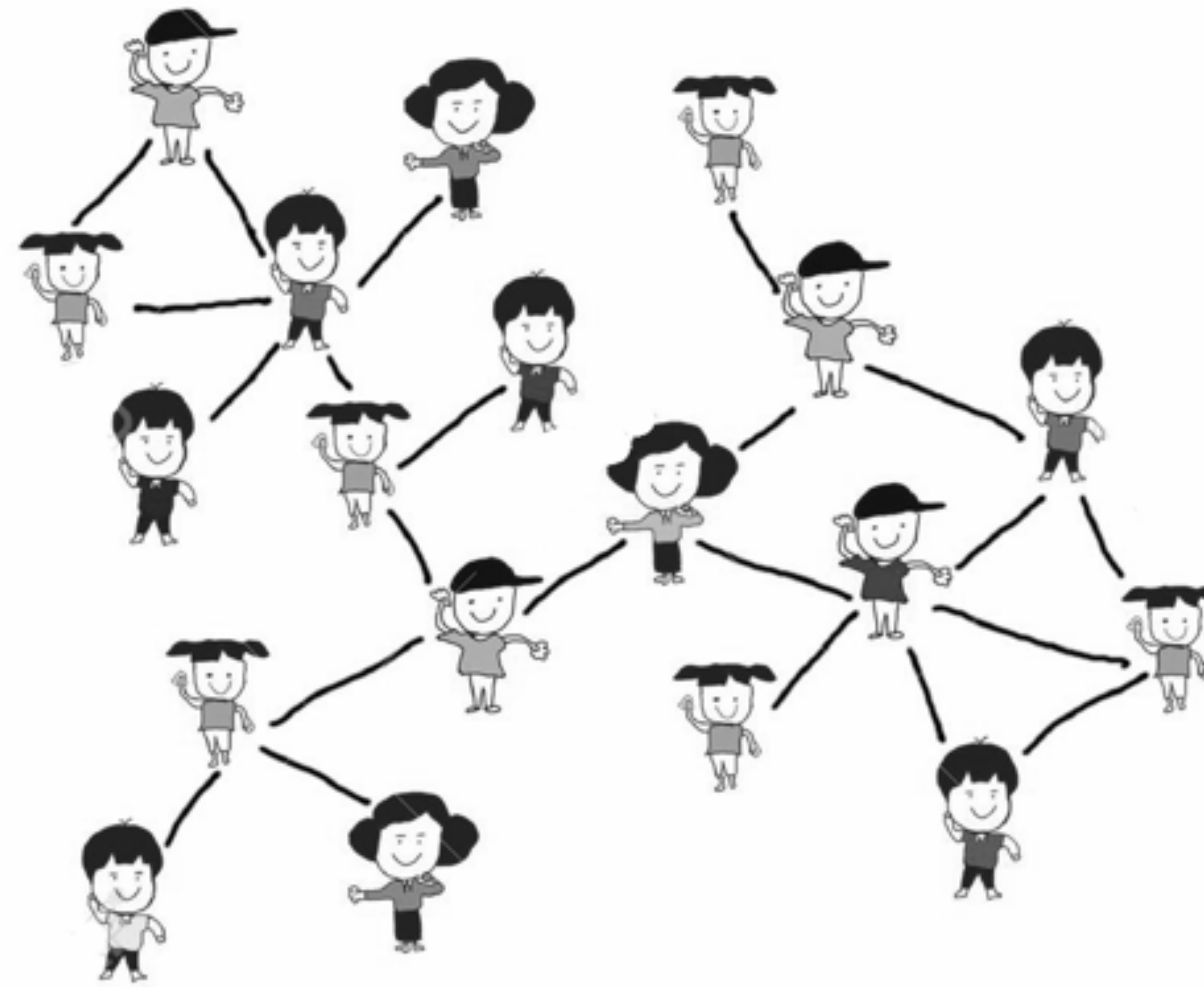
+ Realism
+ Complexity



@gomezgardenes



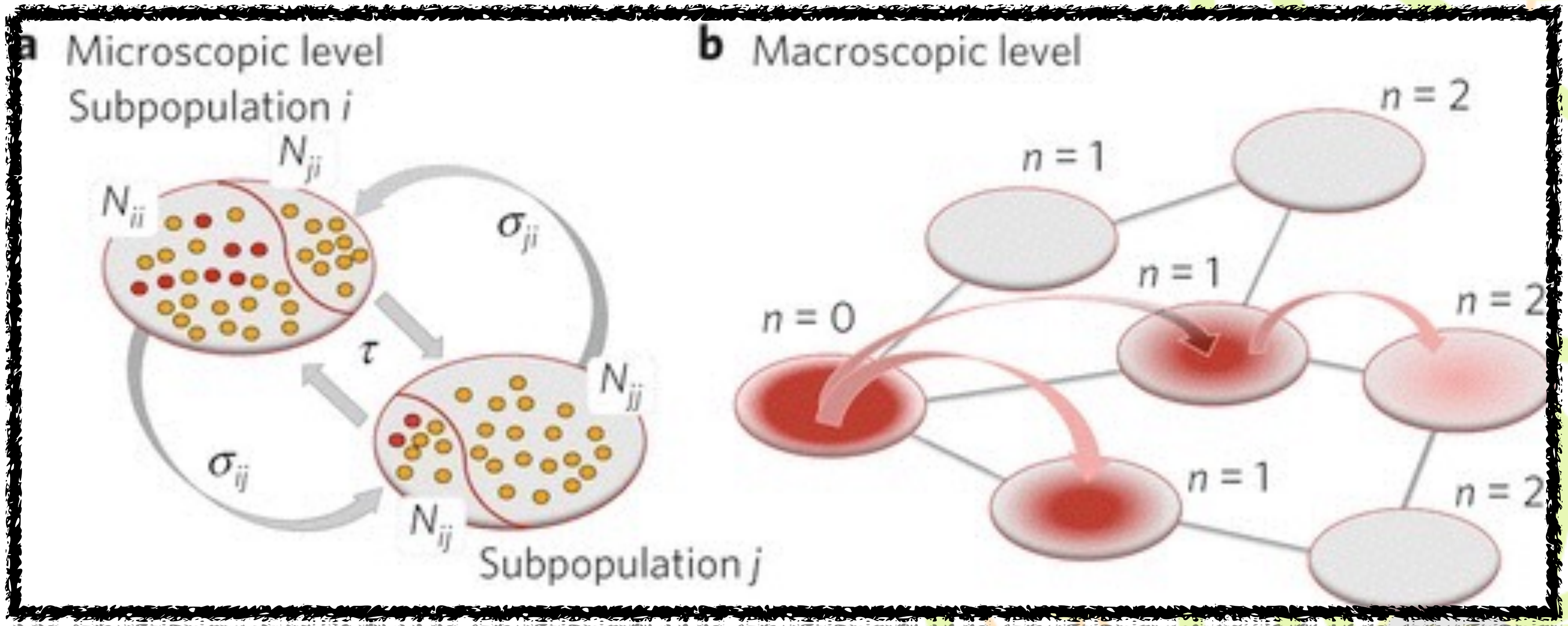
+ Realism
+ Complexity



+ Realism
+ Complexity

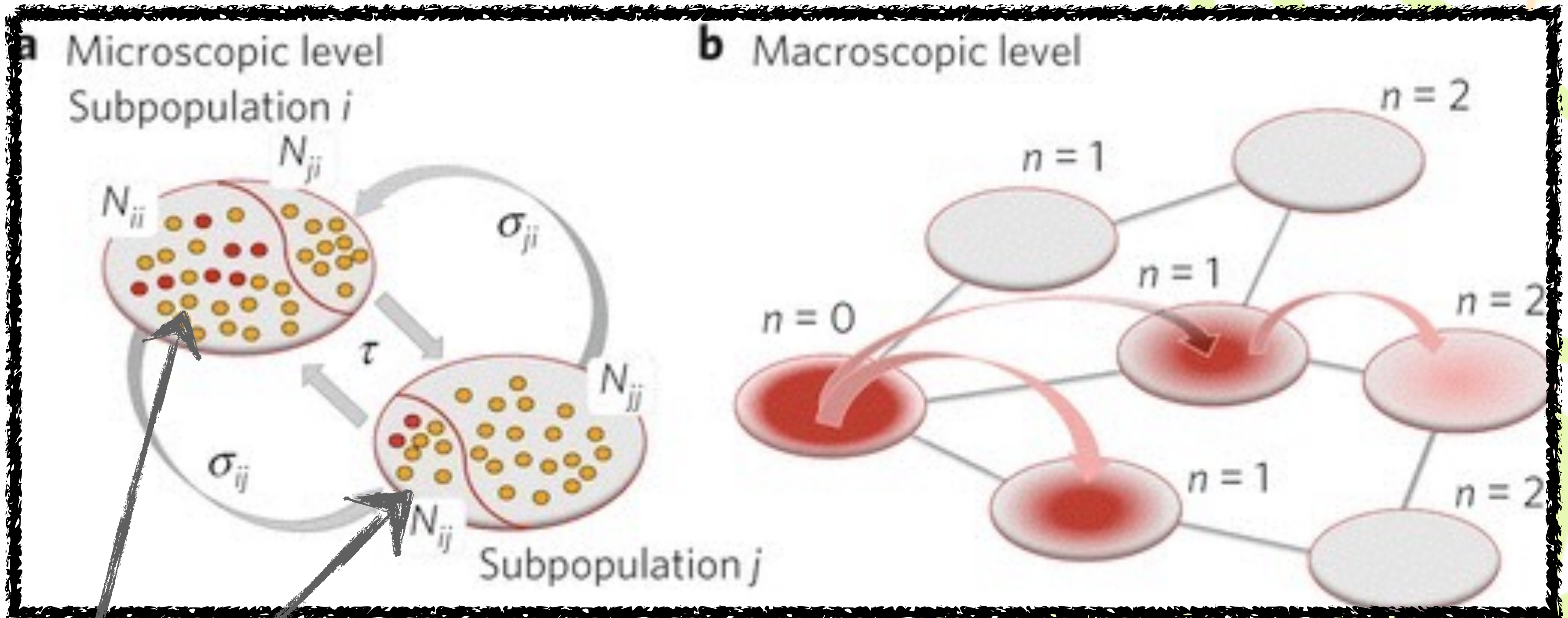
@gomezgardenes

Metapopulation Models



@gomezgardenes

Metapopulation Models

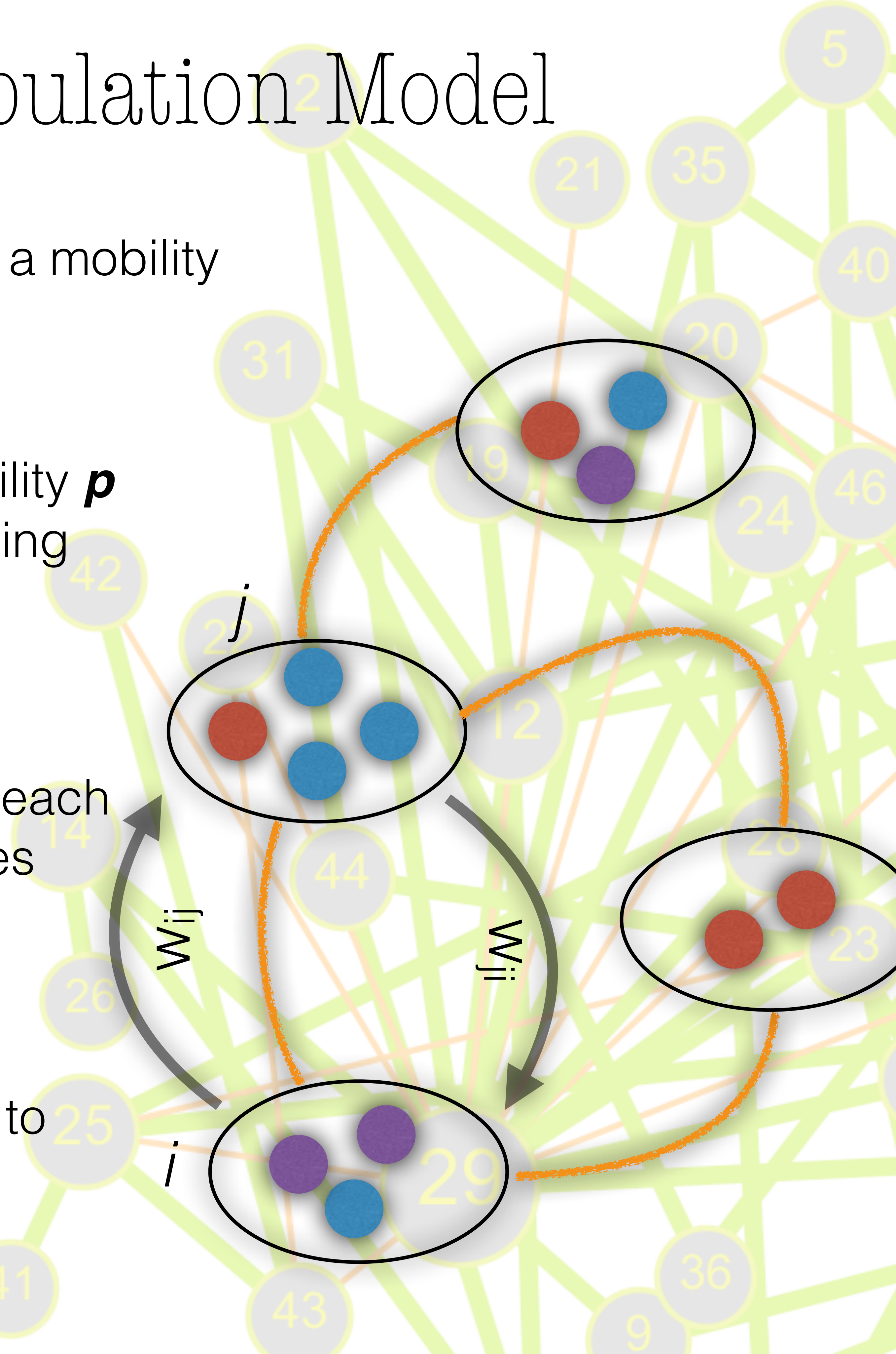


Different levels of description:

- Urban Areas
- Cities
- Regions
- Countries

Basic Metapopulation Model

- Different populations connected by a mobility network (encoded in matrix $\mathbf{W}$).
- Each individual moves with probability ***p*** from one population i to a neighboring one j, according to W_{ij}
- Inside each subpopulation, and at each time step, epidemic dynamics takes place (λ and μ).
- Then, each individual comes back to its original subpopulation (node).



@gomezgardenes

Theory

Available online at www.sciencedirect.com

ScienceDirect

Journal of Theoretical Biology 251 (2008) 450–467

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

ELSEVIER

Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations

Vittoria Colizza[a,*], Alessandro Vespignani[b,c]

PNAS

Multiscale mobility networks and the spatial spreading of infectious diseases

Duygu Balcan[a,b], Vittoria Colizza[c], Bruno Gonçalves[a,b], Hao Hu[d], José J. Ramasco[b], and Alessandro Vespignani[a,b,c,1]

[a]Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408; [b]Pervasive Technology Institute, Indiana University, Bloomington, IN 47404; [c]Computational Epidemiology Laboratory, Institute for Scientific Interchange Foundation, 10133 Torino, Italy; and [d]Department of Physics, Indiana University, Bloomington, IN 47406

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved October 13, 2009 (received for review

Among the realistic ingredients to be considered in the comp
modeling of infectious diseases, human mobility repre
challeng both on the theoretical side and in view o
availability of empirical data. To study the interplay be

ARTICLES

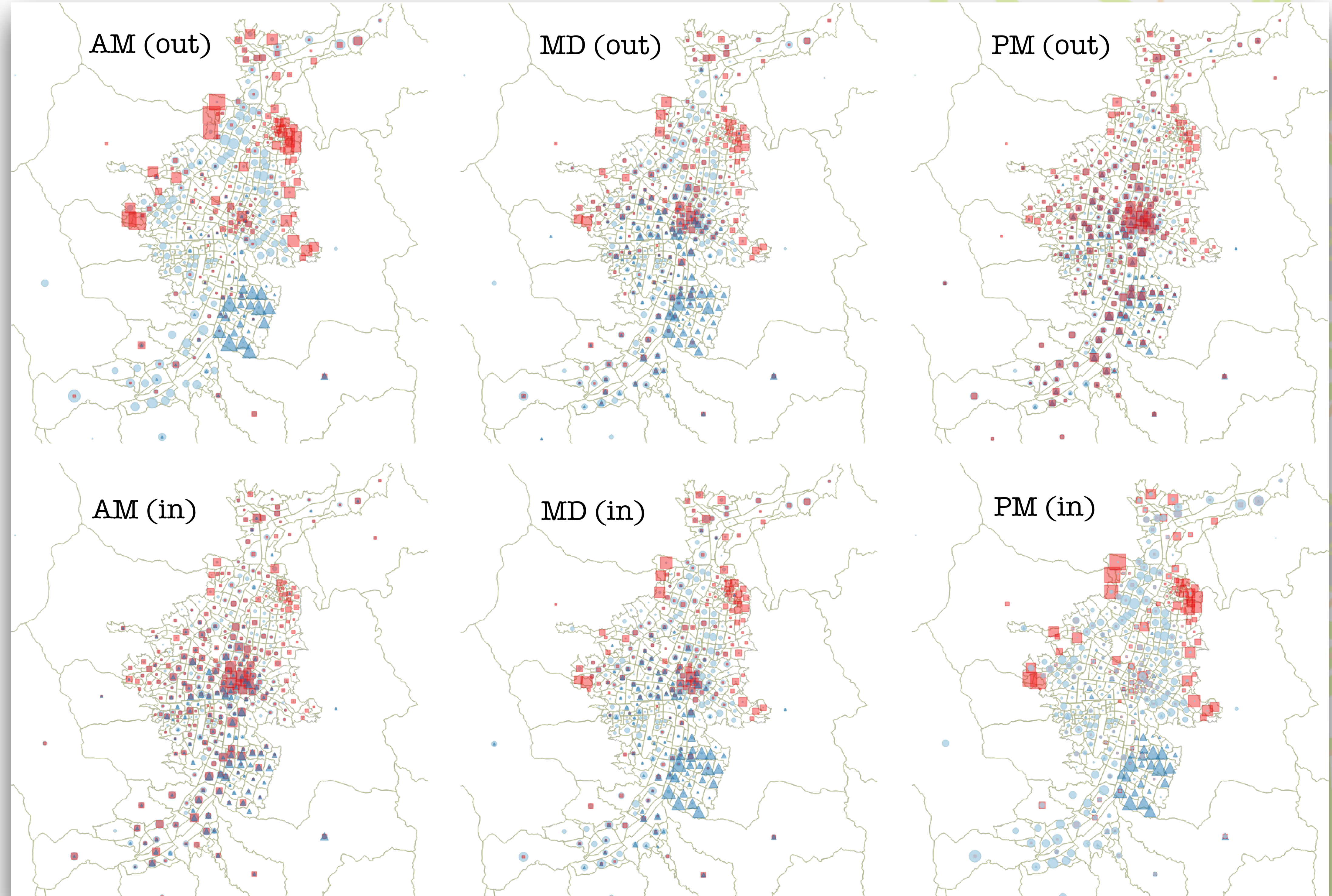
Reaction–diffusion processes and metapopulation models in heterogeneous networks

VITTORIA COLIZZA[1,2*], ROMUALDO PASTOR-SATORRAS[3] AND ALESSANDRO VESPIGNANI[1,2*]

[1]Complex Networks Lagrange Laboratory, Institute for Scientific Interchange (ISI), Torino 10133, Italy
[2]School of Informatics and Department of Physics, Indiana University, Bloomington 47406 Indiana, USA
[3]Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain

@gomezgardenes

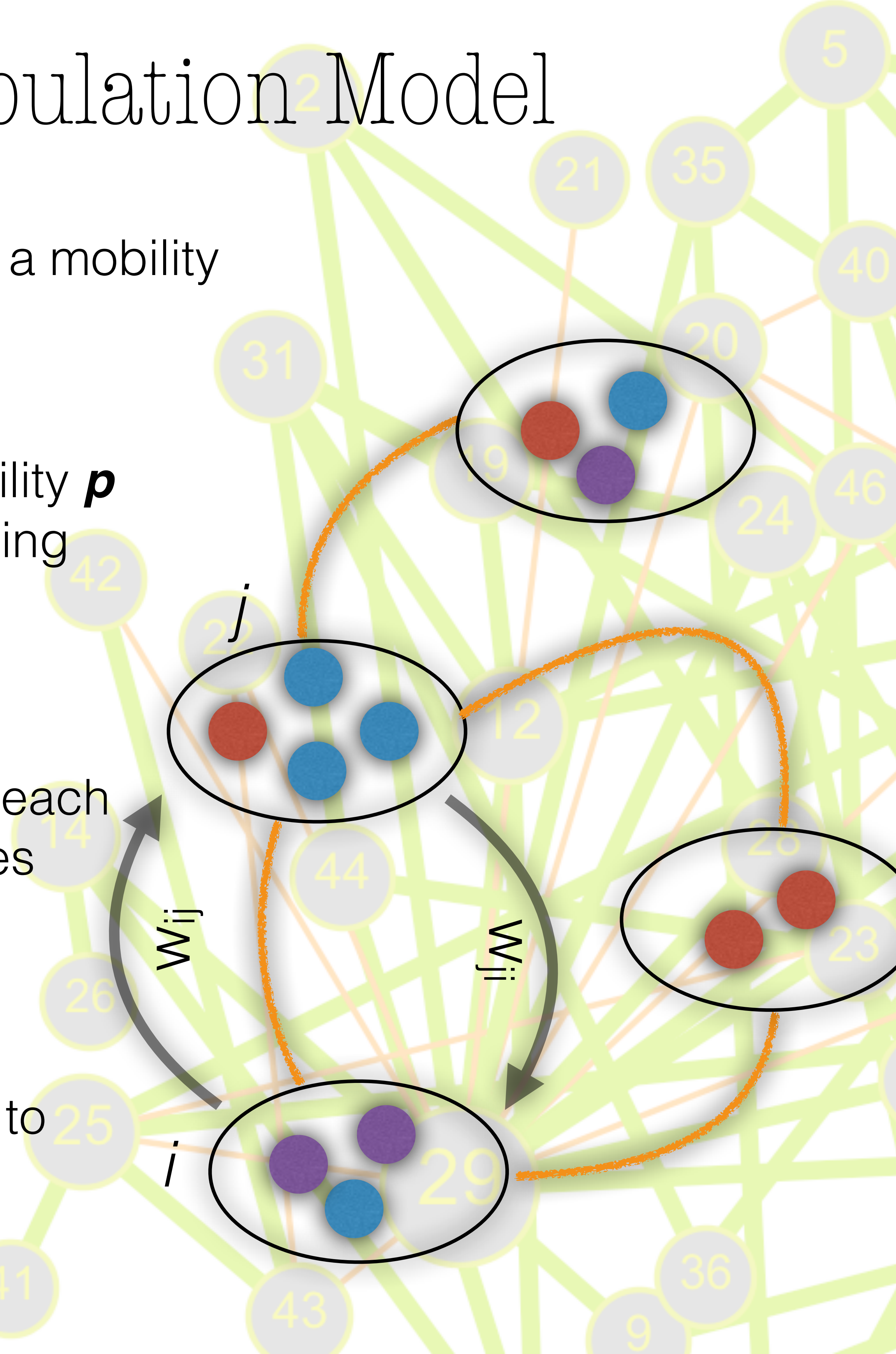
Real mobility patterns



@gomezgardenes

Basic Metapopulation Model

- Different populations connected by a mobility network (encoded in matrix $\mathbf{W}$).
- Each individual moves with probability ***p*** from one population i to a neighboring one j, according to W_{ij}
- Inside each subpopulation, and at each time step, epidemic dynamics takes place (λ and μ).
- Then, each individual comes back to its original subpopulation (node).



@gomezgardenes

Basic Metapopulation Model

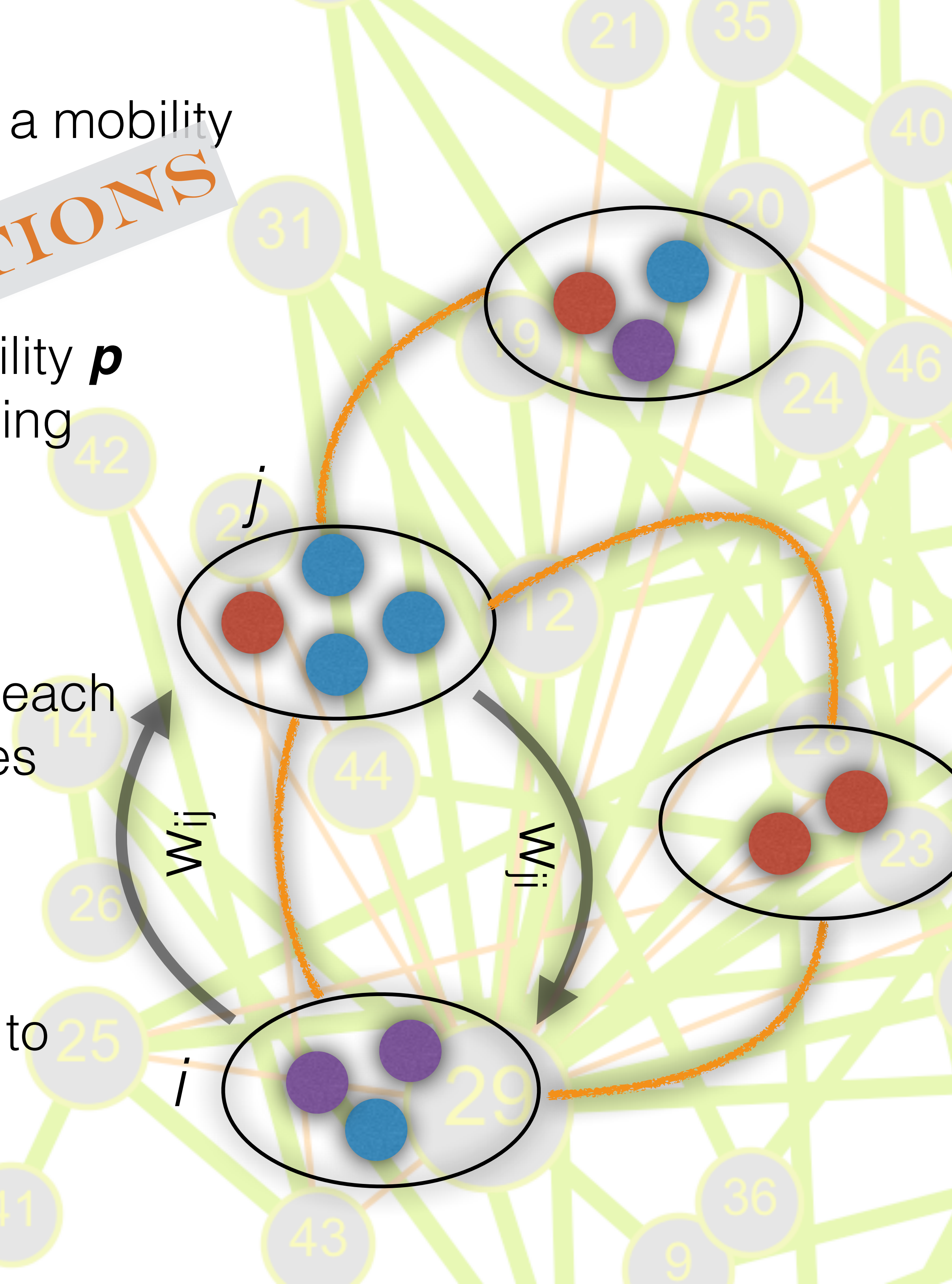
- Different populations connected by a mobility network $\boxed{\text{(encoded in matrix } \mathrm{W})}$.

EMPIRICAL OBSERVATIONS

- Each individual moves with probability ***p*** from one population i to a neighboring one j, according to $\boxed{W_{ij}}$

- Inside each subpopulation, and at each time step, epidemic dynamics takes place (λ and μ).

- Then, each individual comes back to its original subpopulation (node).



@gomezgardenes

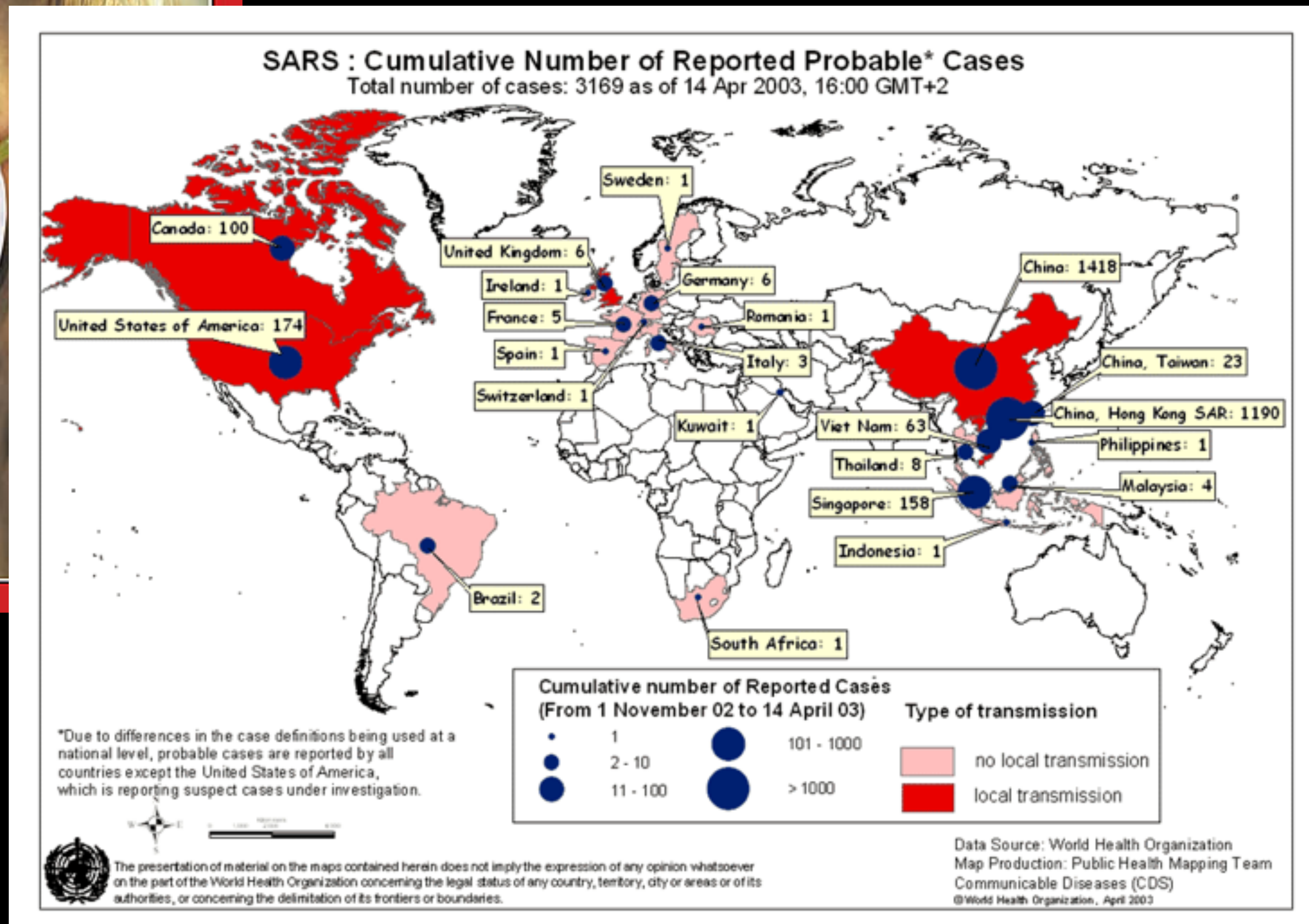
A case study:

SARS Pandemic 2003

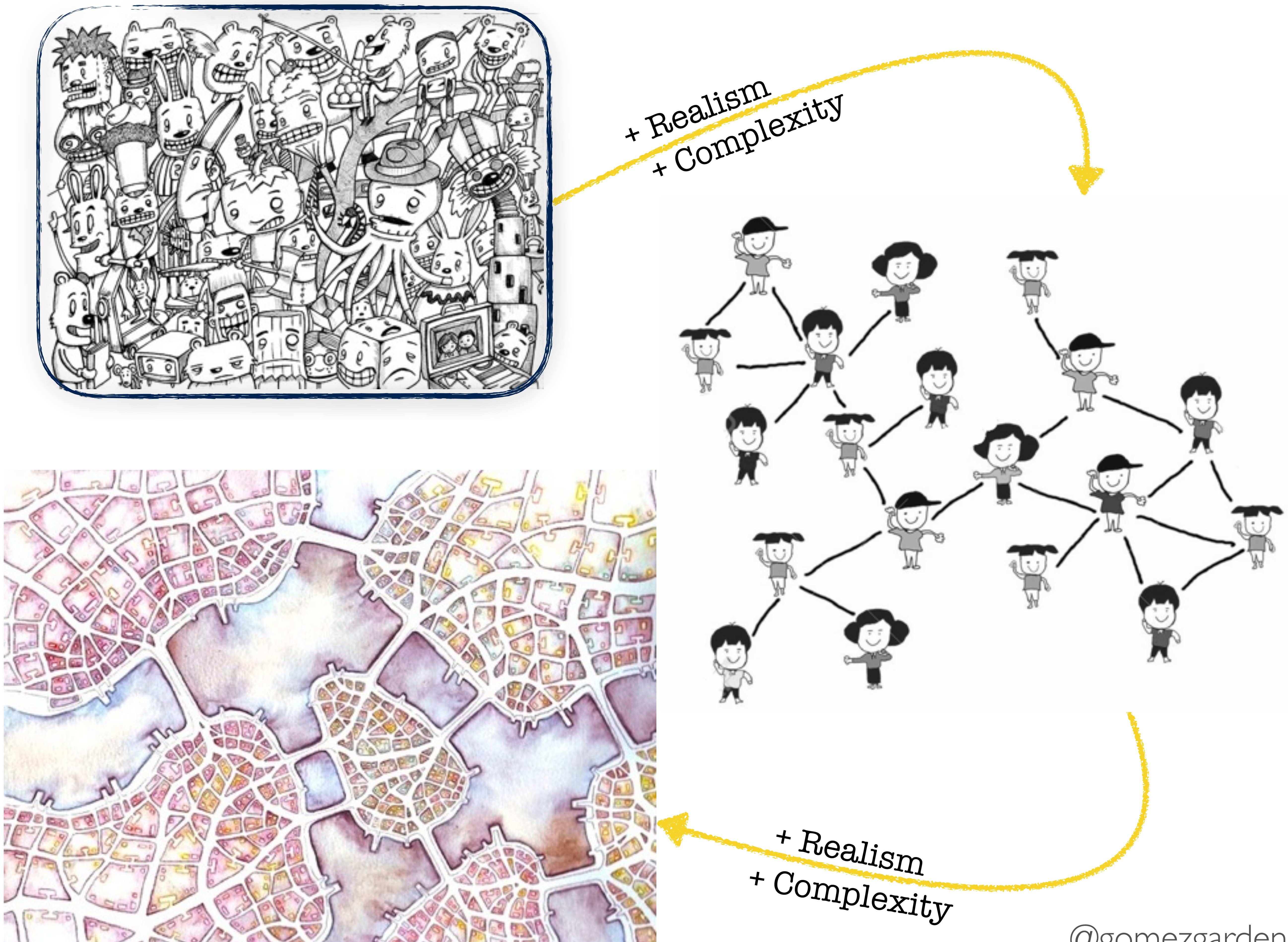
@gomezgardenes

A case study:

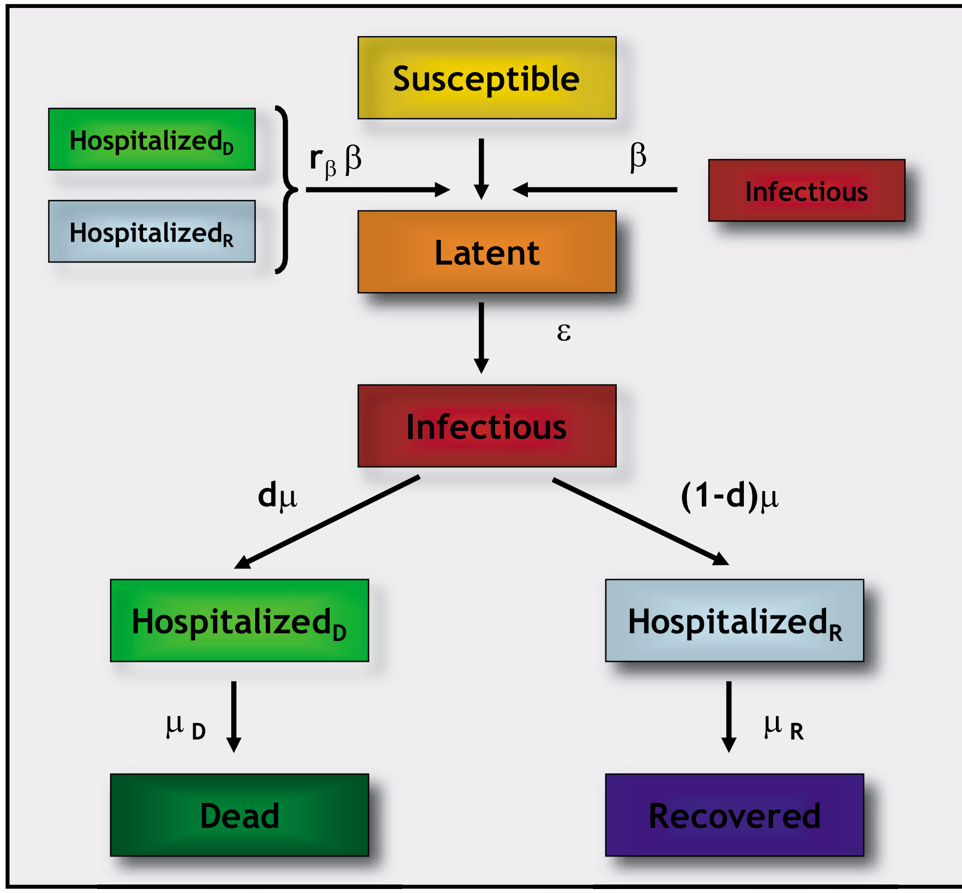
SARS Pandemic 2003



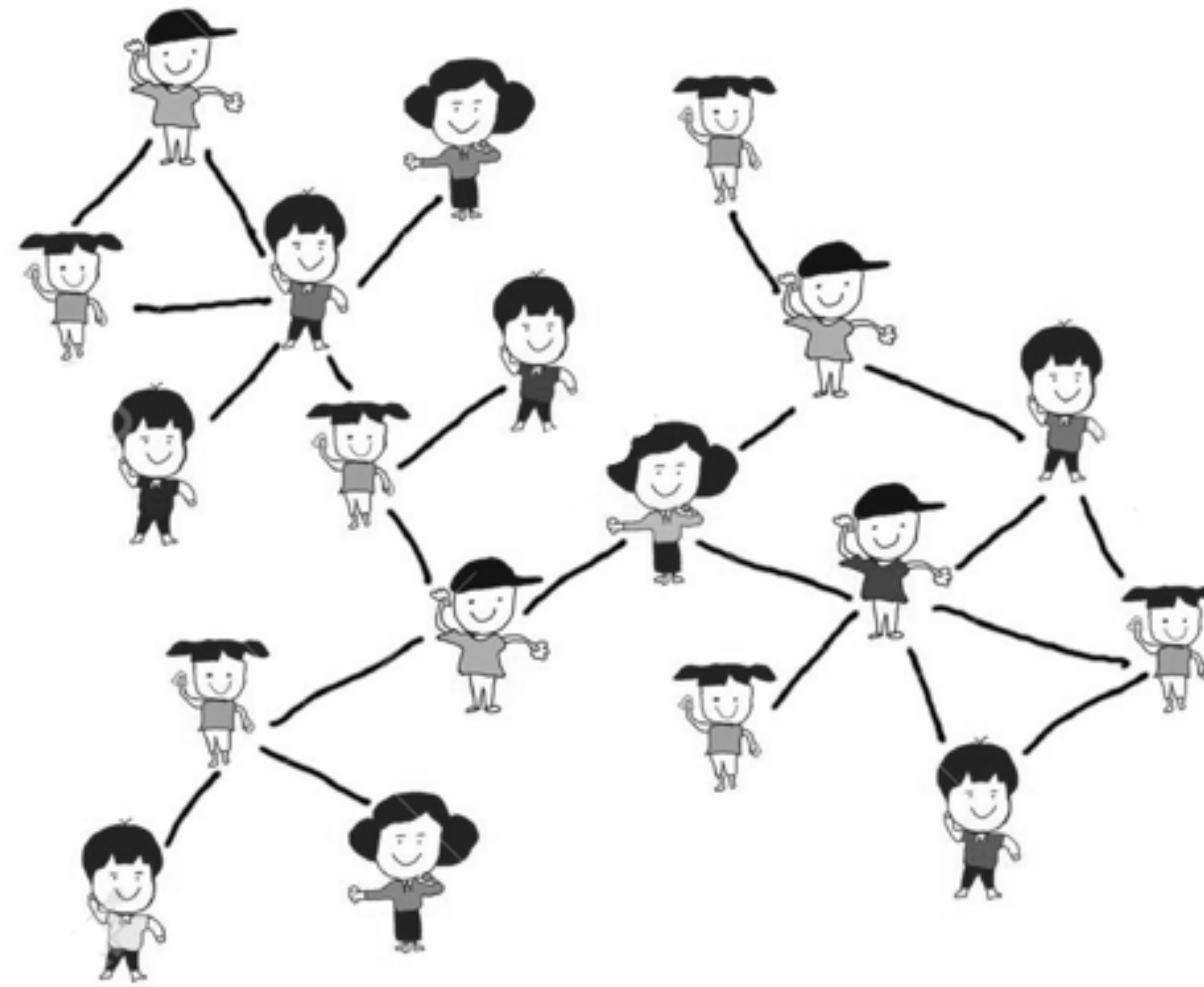
@gomezgardenes

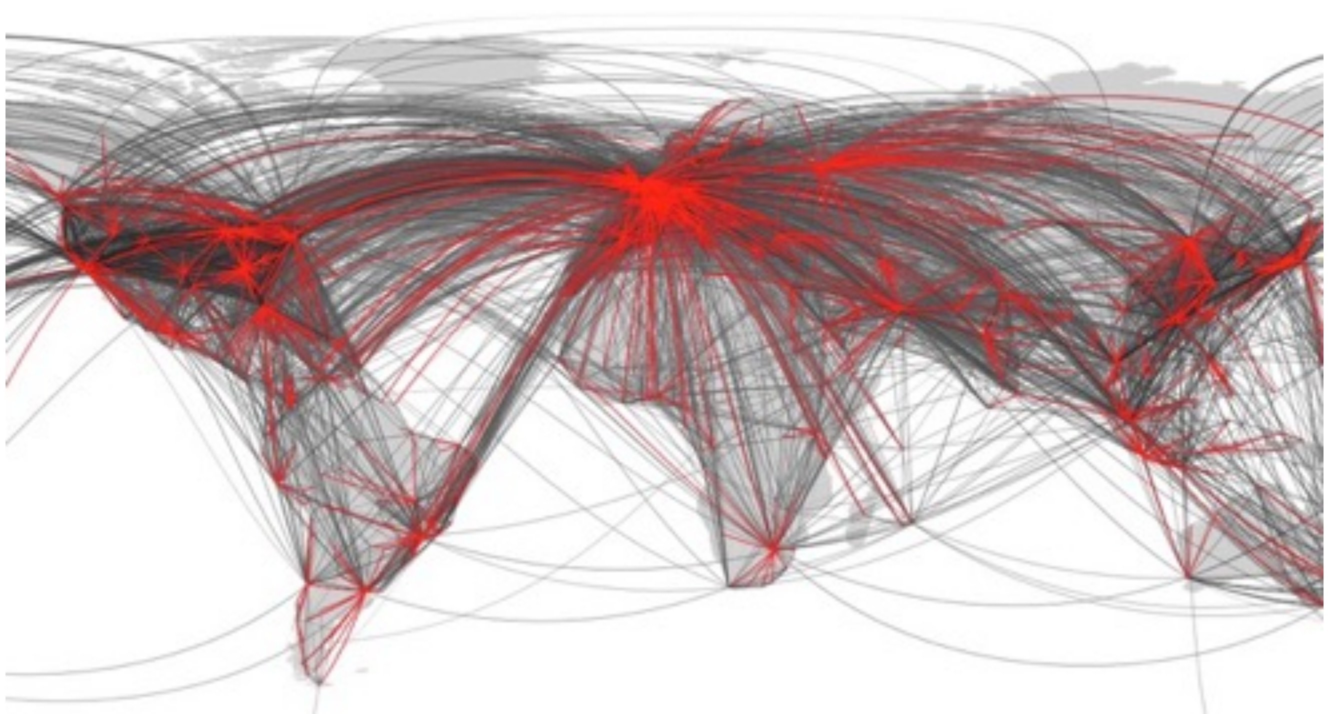


@gomezgardenes



+ Realism
+ Complexity





IATA DATABASE + URBAN CENSUS

+ Realism
+ Complexity

@gomezgardenes

Model Validation

BMC Medicine BioMed Central

Research article **Open Access**

Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study

Vittoria Colizza*[1], Alain Barrat[1,2], Marc Barthélemy[3] and Alessandro Vespignani[4,5]

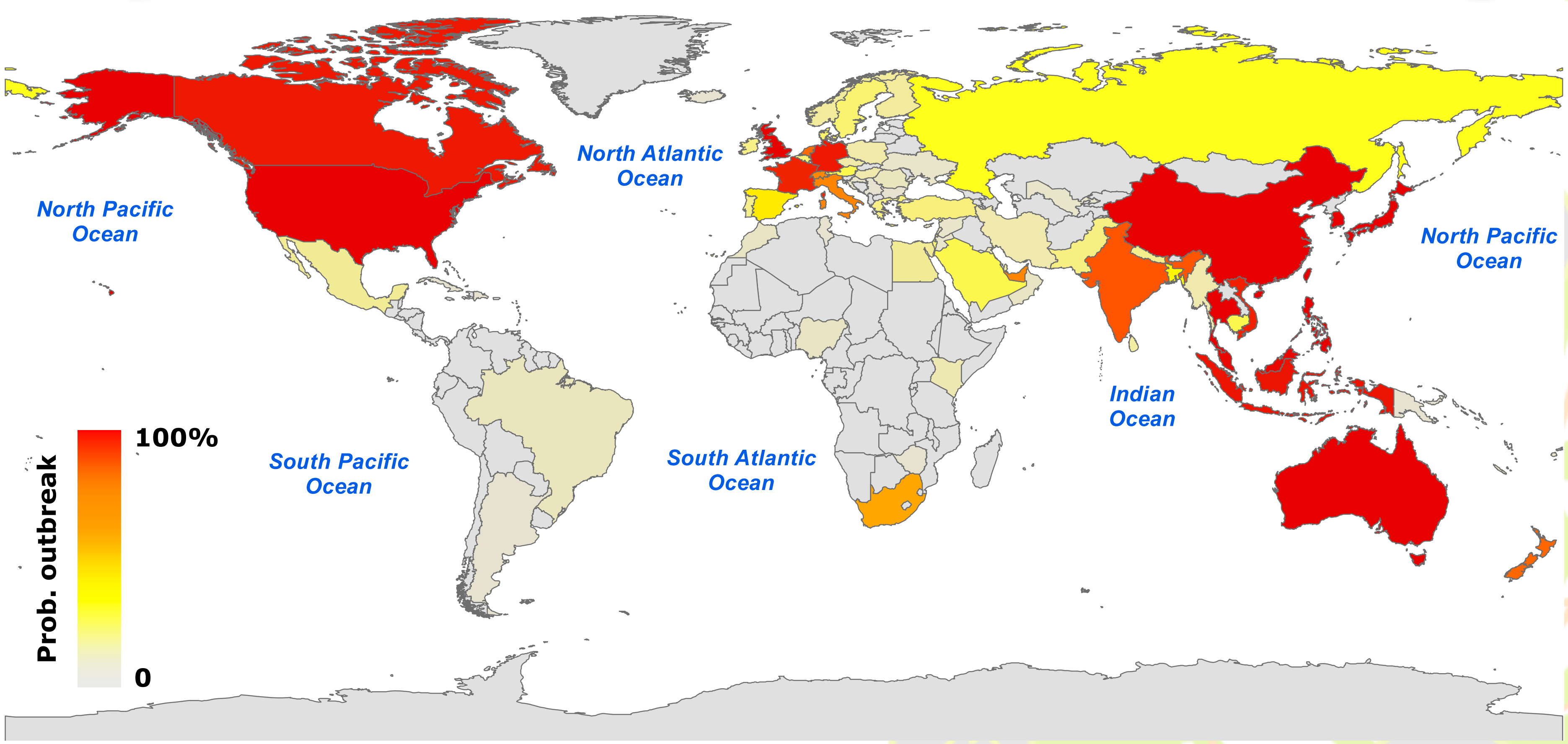


Figure 2
Worldwide map representation of the outbreak likelihood as predicted by the stochastic model. Countries are represented according to the color code, ranging from gray for low outbreak probability to red for high outbreak probability.

@gomezgardenes

Model Validation

BMC Medicine

BioMed Central

Research article

Open Access

Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study

Vittoria Colizza*[1], Alain Barrat[1,2], Marc Barthélemy[3] and Alessandro Vespignani[4,5]

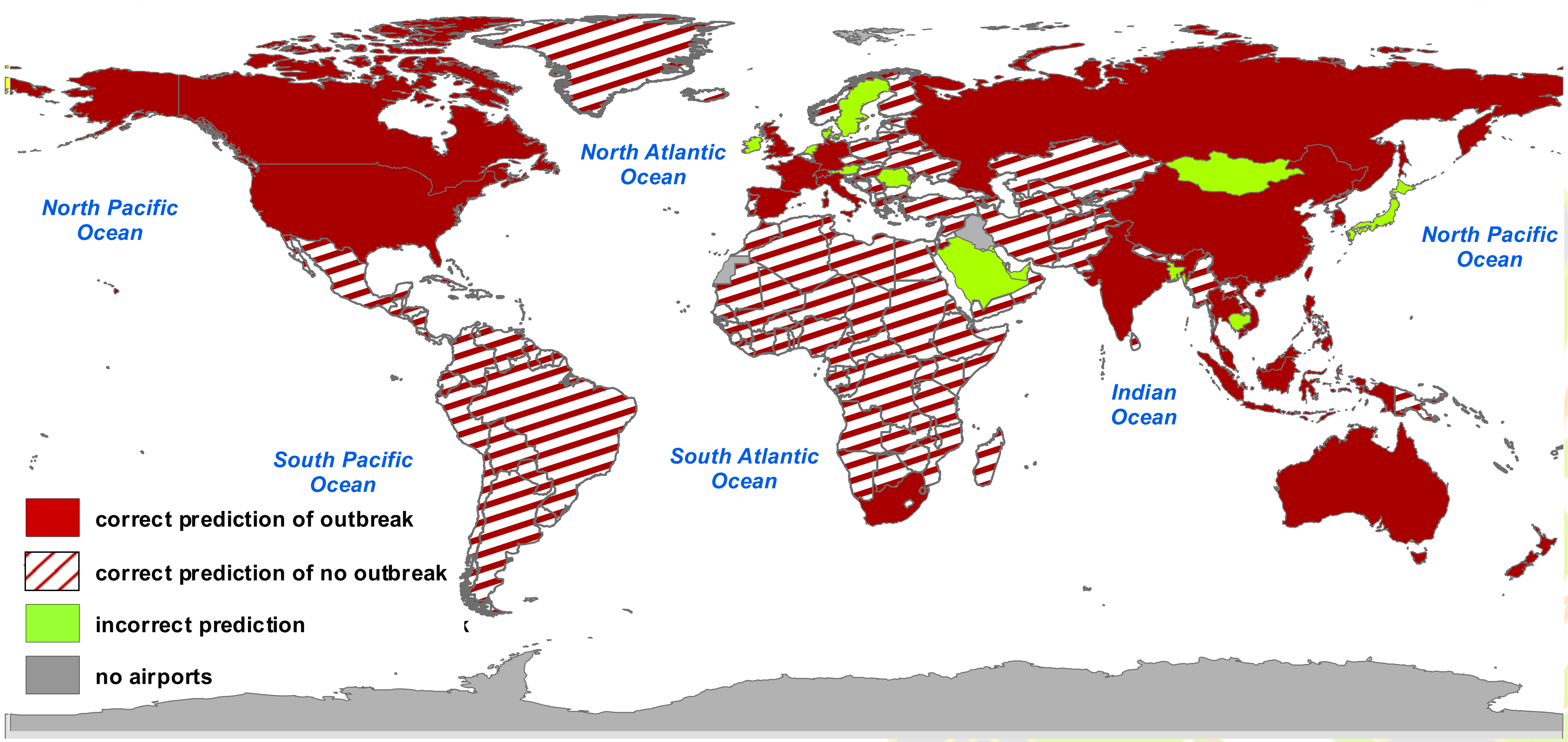


Figure 4

Map representation of the comparison between numerical results and WHO reported cases. Countries are considered at risk if the probability of reporting an outbreak – computed on $n = 10^3$ different realizations of the stochastic noise – is larger than 20%. In red we represent countries for which model forecasts are in agreement with WHO official reports, distinguishing between correct predictions of outbreak (filled red) and correct predictions of no outbreak (striped red). Forecasts that deviate from observed data are represented in green. Results shown refer to the date of 11 July 2003.

@gomezgardenes

Model Validation

BMC Medicine

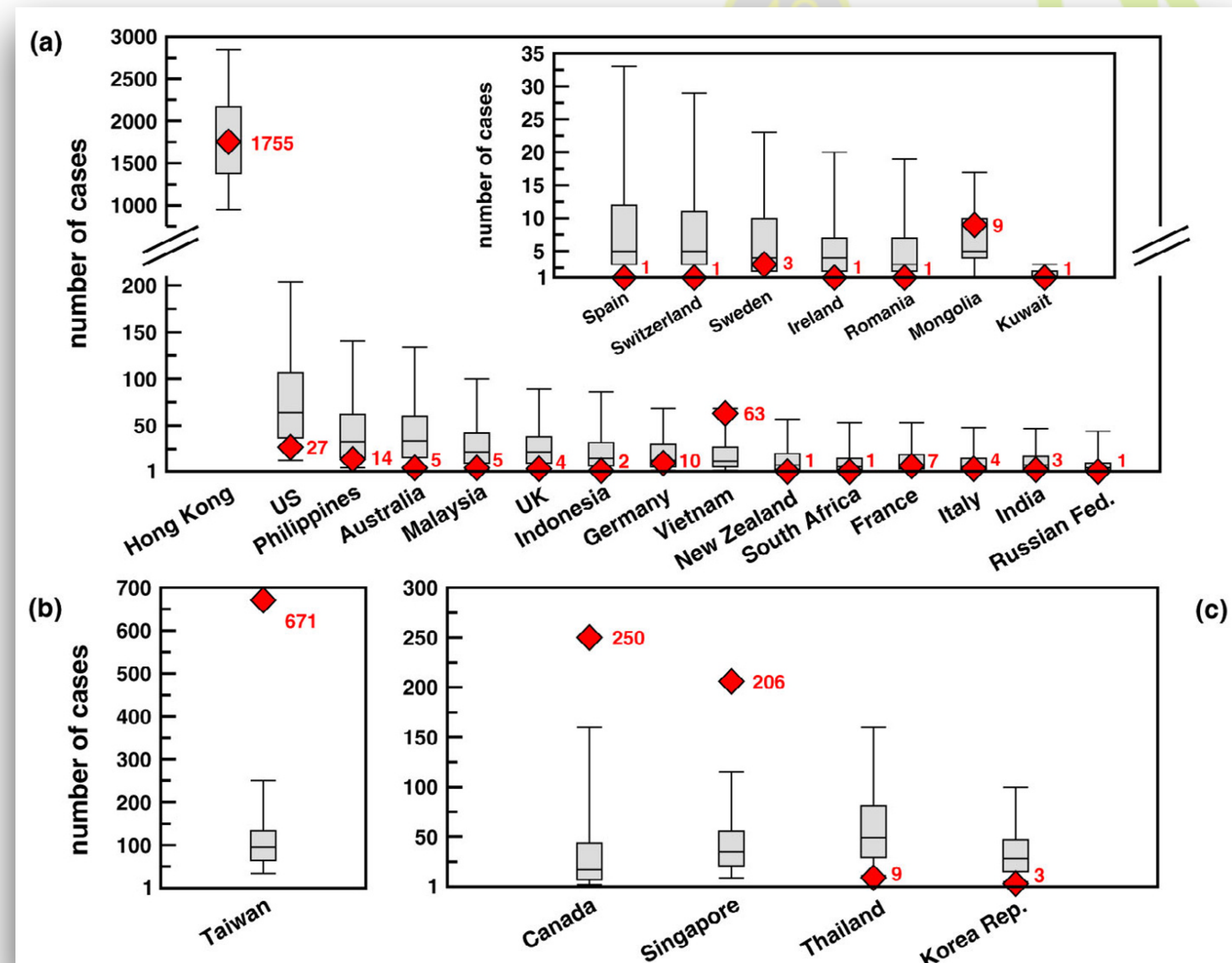
BioMed Central

Research article

Open Access

Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study

Vittoria Colizza*[1], Alain Barrat[1,2], Marc Barthélemy[3] and Alessandro Vespignani[4,5]



@gomezgardenes

Model Validation

BMC Medicine

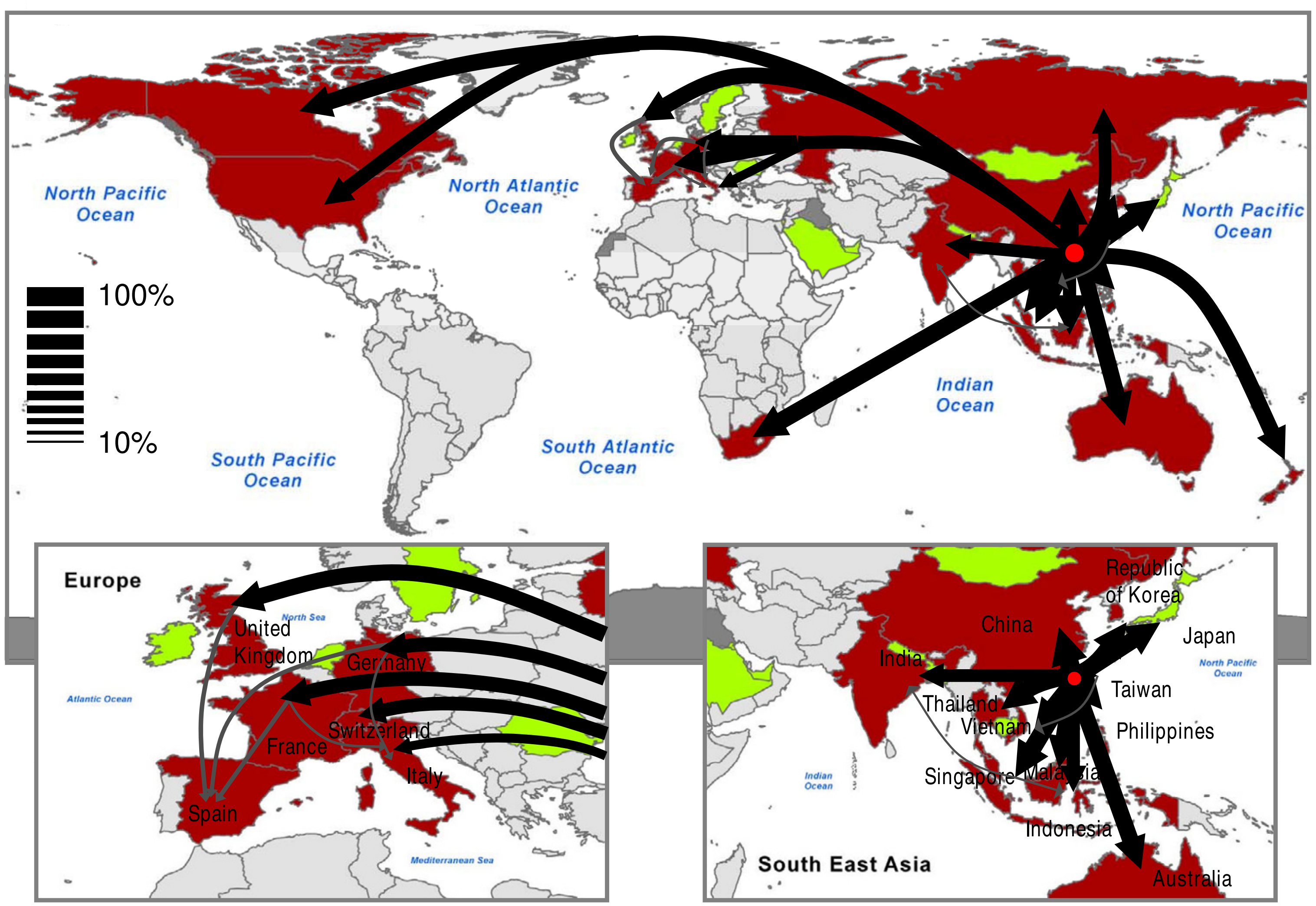
BioMed Central

Research article

Open Access

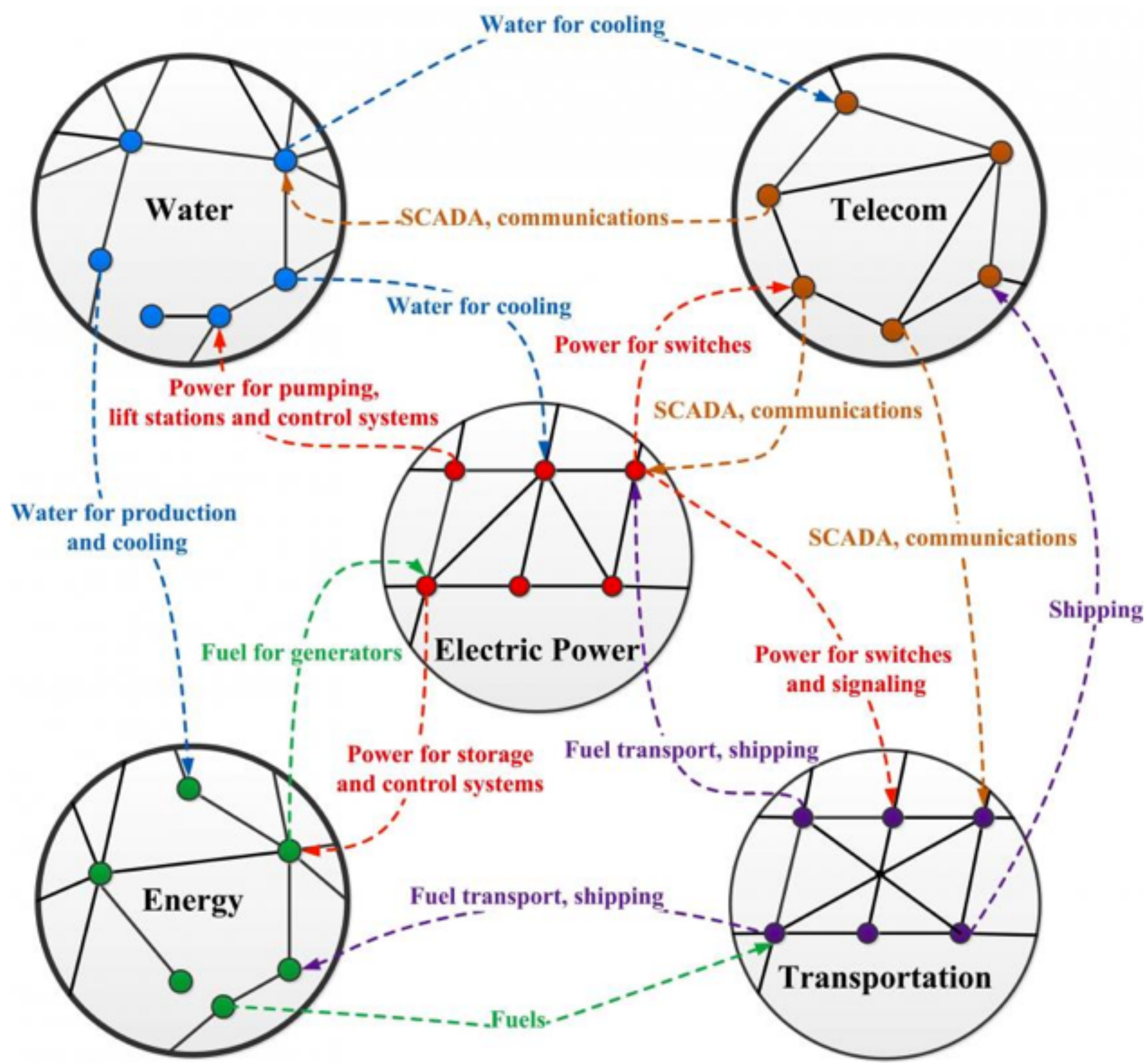
Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study

Vittoria Colizza*[1], Alain Barrat[1,2], Marc Barthélemy[3] and Alessandro Vespignani[4,5]



@gomezgardenes

EPIDEMIC FORECAST?



Ginestra's Lecture: Multilayer Networks

@gomezgardenes

Thanks for your attention

The lecture is available at my website:

http://complex.unizar.es/~jesus/

@gomezgardenes