Metric Spaces in Bipartite Systems

Maksim Kitsak, Dmitri Krioukov, Fragkiskos Papadopoulos, Marian Boguñá, M. Ángeles Serrano and kc claffy

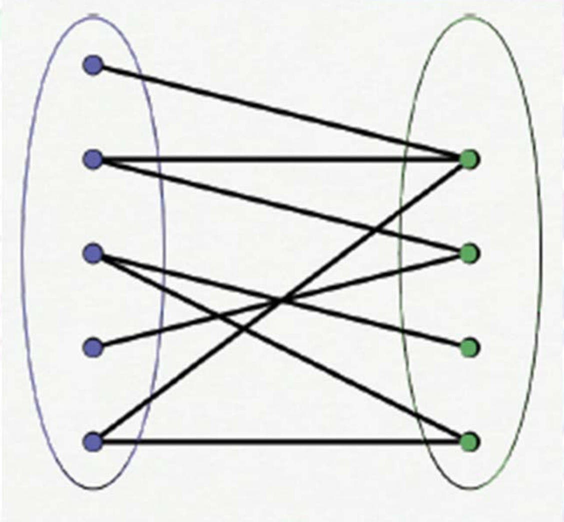
(1) High Bipartite Clustering as a Consequence of the Triangle Inequality.

(2) Class of Bipartite Networks in Metric Spaces

(3) Scaling of Bipartite Clustering

CCNR, Boston
May,17,2010

What is a bipartite network? Definition and Examples



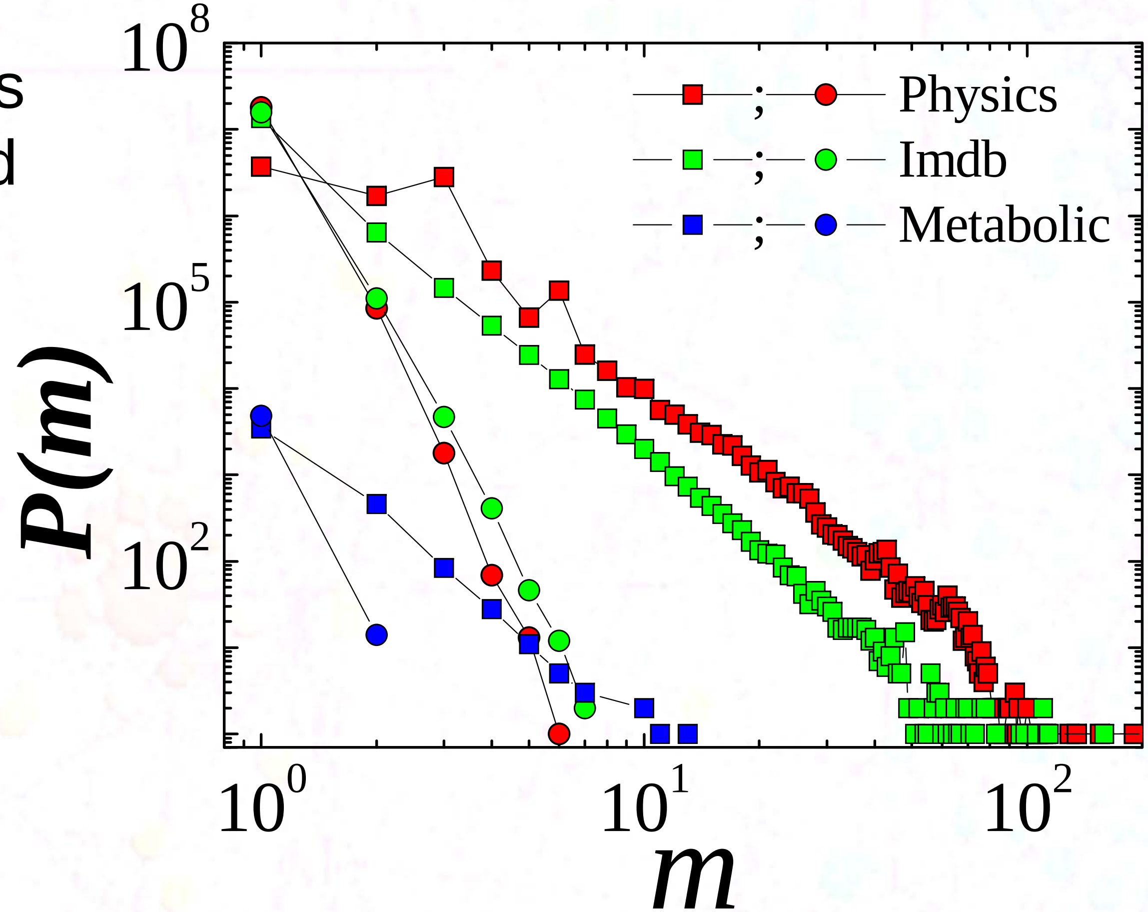
Nodes of a bipartite network can be divided into two disjoint sets (authors, papers) so that no links connect 2 nodes in the same set.

Examples:

- Collaboration networks:
  Authors are associated with papers they publish
- Actor networks:
  Actors are connected to films.
- Metabolic Networks
  Metabolites are related to chemical reactions
- Peer to peer networks (P2P):
  Participants that make a portion of their resources directly available to other network participants.

How many papers two authors have in common?

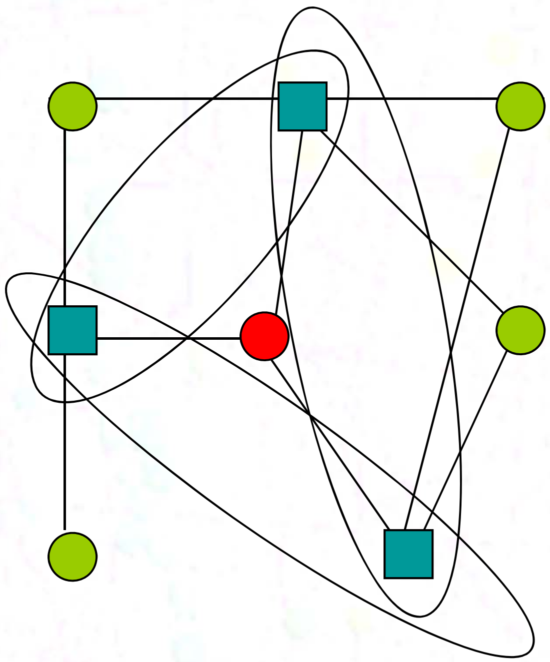
(**m**): # of shared neighbors
(**M**): maximum # of shared neighbors
Physics: M=190(6)
Imdb: M=216 (7)
Metabolic: M=13 (2)



1) P(m) is distributed as a power-law.
2) M is significantly higher in real bipartite networks than in randomized.

Bipartite Clustering Coefficient

1) Bipartite Networks: Neighbors of a given node are **NEVER** connected. C=0. No 3-loops in bipartite networks.
2) Bipartite Clustering is defined based on 4-loops!



Consider all pairs of neighbors!

1 node in common, **4** nodes altogether

0 nodes in common, **4** nodes altogether.

2 Nodes in common, **3** nodes altogether.

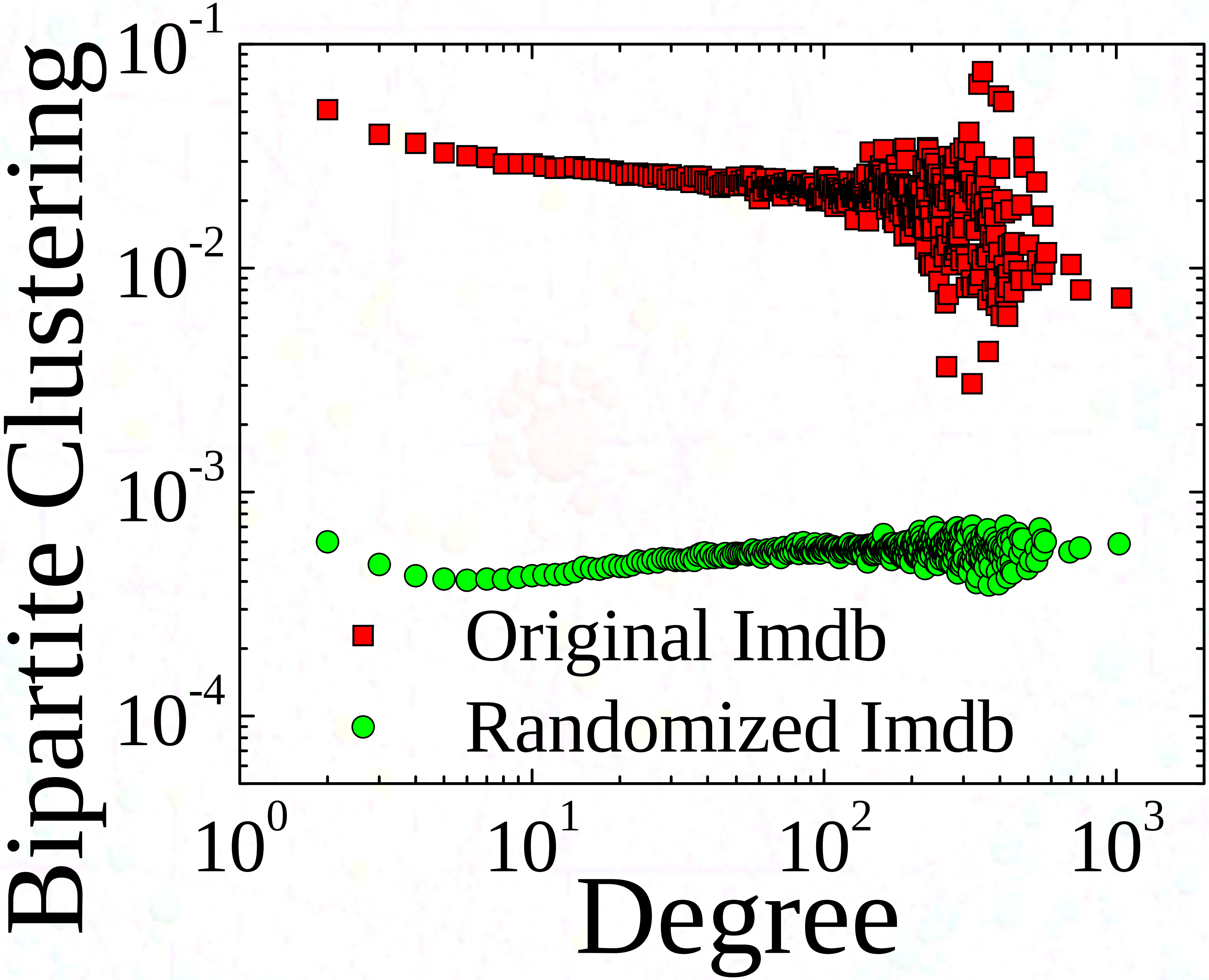
$$C = \frac{1+0+2}{4+4+3} = \frac{3}{11}$$

$$C = \frac{\sum_{m \neq n} A_m \cap A_n}{\sum_{m \neq n} A_m \cup A_n}$$

*P. Zhang et al, Physica A, **387** 27 6869 (2008).*

Bipartite clustering is significantly higher in real bipartite networks than in random networks (Next Slide).

Real bipartite networks are highly clustered



Properties of Bipartite Networks

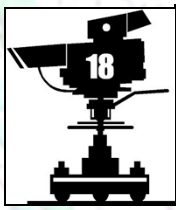
1) Top (Bottom) nodes tend to share a lot of (Bottom) (Top) nodes.
2) Bipartite networks are highly clustered.

WHY?

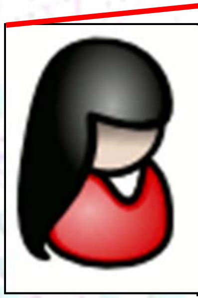
Bipartite networks have metric structure.

Actor network has metric structure?

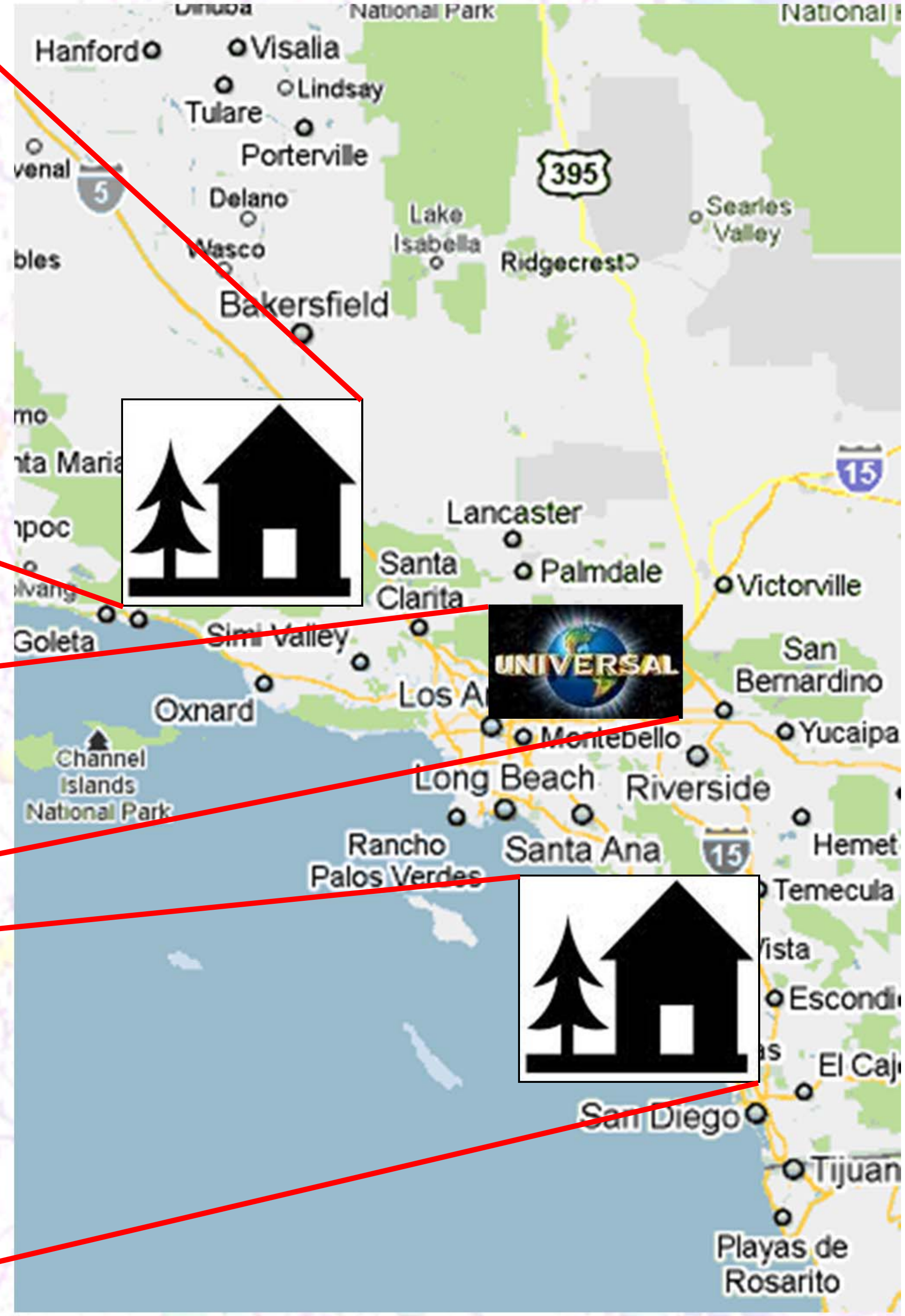
Name: ***Mary***
Age: ***21***
Comedies: ***5***
Santa Barbara, CA



..another teenager comedy
20 female actresses
20 male actors



Name: ***Jessica***
Age: ***22***
Comedies: ***8***
Oceanside, CA



Actor network has metric structure?

A. *smaller distances imply higher connection probability!*

B. *small* $\mathbf{d_1}$ *and* $\mathbf{d_2}$ *imply small* $\mathbf{d_3}$*?*

C. *Triangle Inequality* $d_3 \leq d_1 + d_2$*?*

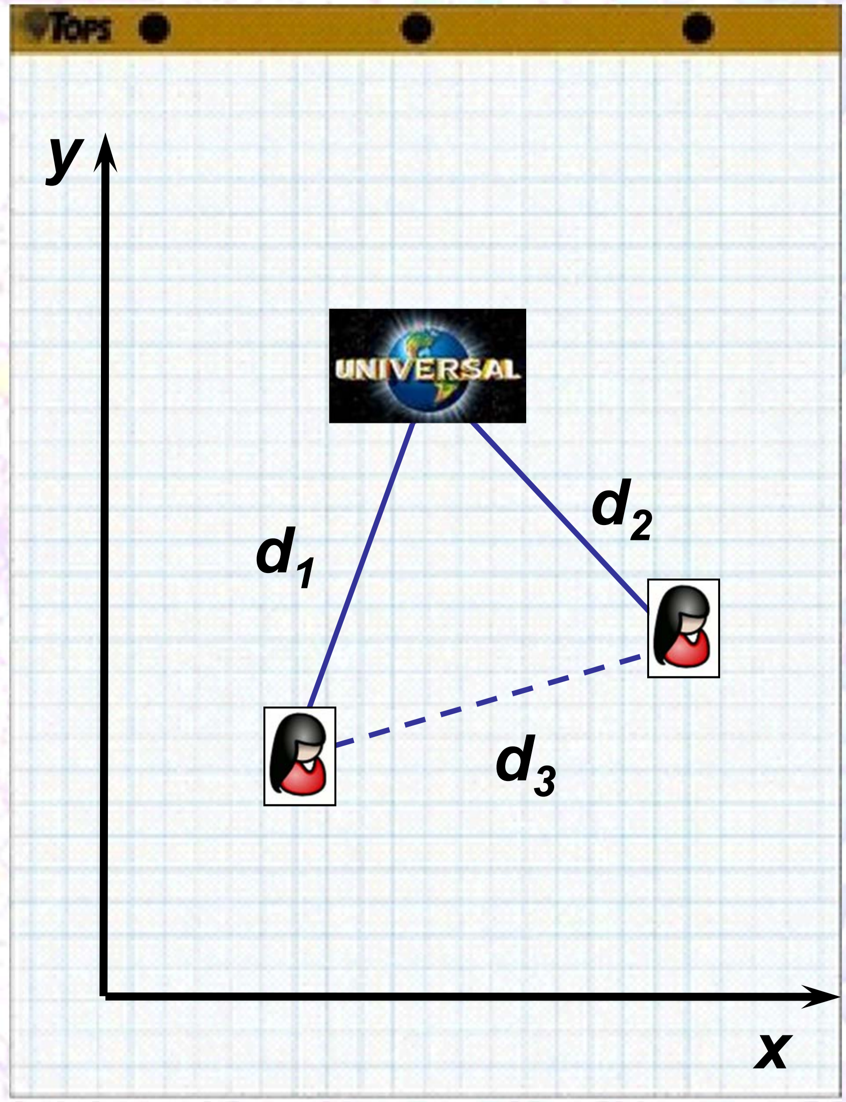
Underlying Space is Metric!

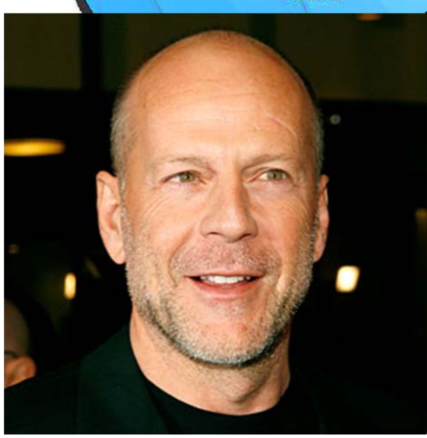
1) $d(x,y) \geq 0$

2) $d(x,y)=0 \leftrightarrow x=y$

3) $d(x,y)=d(y,x)$

4) $d(x,z) \leq d(x,y)+d(z,y)$





Name: ***Bruce Willis***

Age: ***55***

Movies: ***82***

Genres: ***Comedy, Drama, Action, Thriller, Romance, Sci-Fi…***

Los Angeles, CA

Higher Degree nodes are likely to connect at large distances!

The Underlying Metric Space Hypothesis

A. *Top and Bottom nodes of bipartite networks exist in underlying* ***metric*** *spaces.* ($d_3 \leq d_1 + d_2$)

B. *The probability of a link connecting a pair of nodes is determined by the geometric distance between the nodes in the underlying space.*

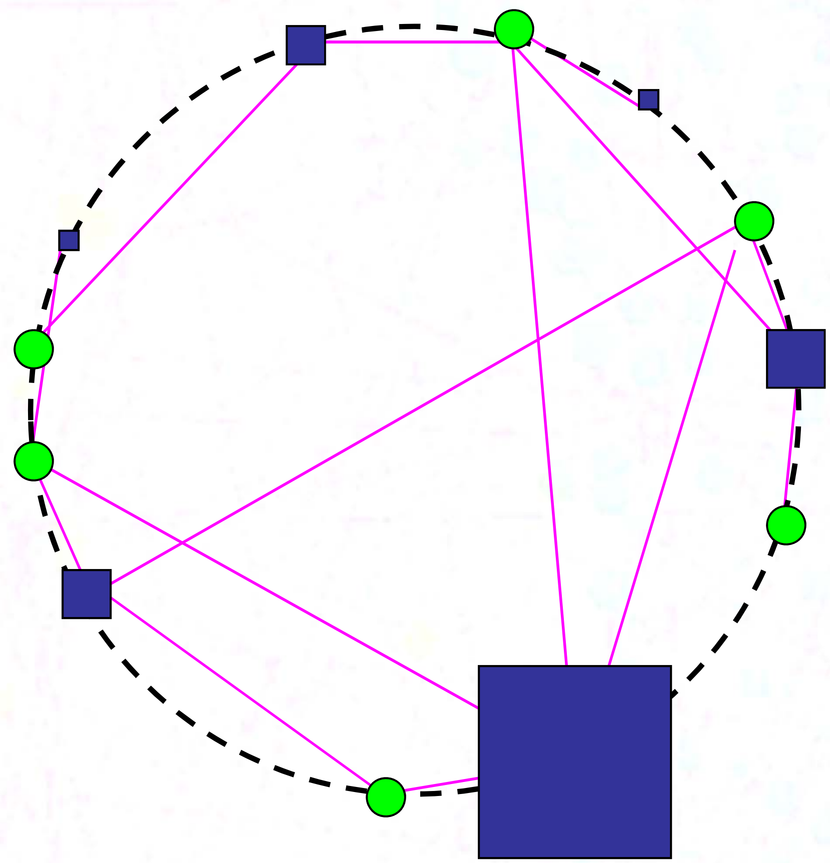
C. *The probability of a link is specified by a connection probability function* $r(d/d_c)$. $r(x)$ *can be any decreasing function of* x.

D. *Every node is assigned an intrinsic fitness parameter:* κ *(top nodes)* λ *(bottom nodes).*

E. *In a Euclidean D-dimensional space the characteristic scale is* $d_c \sim [\kappa\lambda]^{1/D}$.

Modeling Bipartite Networks in Metric Spaces: S1S1 model

1) Uniformly distribute ***N*** top and ***M*** bottom nodes on a 1-D Euclidean ring of radius ***R***.



2) For every top (bottom) node calculate its fitness **κ** *(**λ**) drawn from a pdf **P(κ)** (**P(λ)**).*

$$P(\kappa) \sim \kappa^{-\gamma_1}$$

$$P(\lambda) \sim \lambda^{-\gamma_2}; P(\lambda) = \delta(\lambda - \lambda_0)$$

3) Connect authors and papers according to the connection probability function:

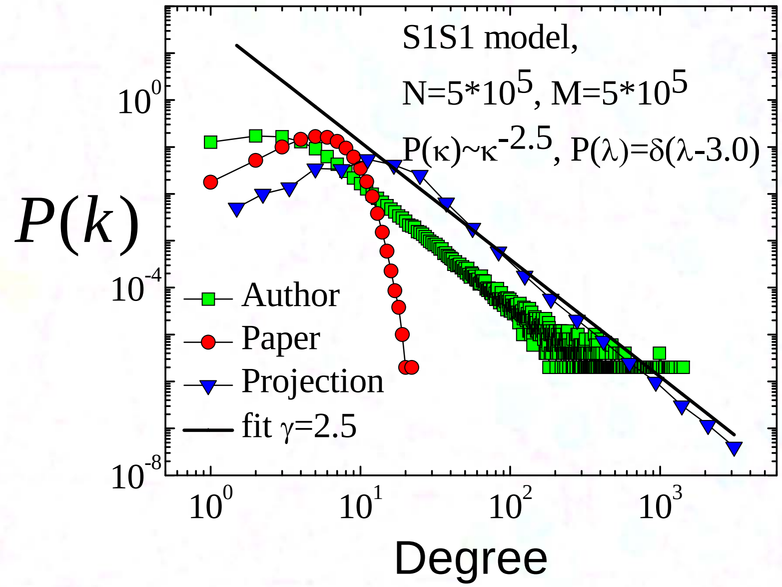
$$p_{i,j} = \frac{1}{1 + \left(\frac{d_{i,j}}{\mu \kappa_i \lambda_j} \right)^{\beta}}$$

Connectivity of S1S1 model.

1) Resulting degrees are proportional to node magnitudes:

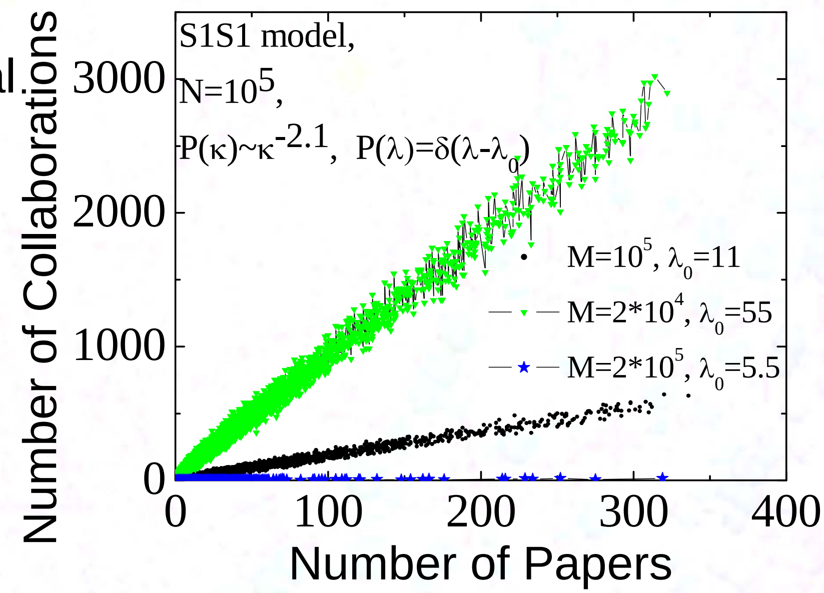
$$\bar{k}_{author}(\kappa) \propto \kappa \qquad P_{author}(k) \sim k^{-\gamma}$$

$$\bar{k}_{paper}(\lambda) \propto \lambda \qquad P_{paper}(k) \sim e^{-\lambda_0} \lambda_0^{\,k} / k!$$



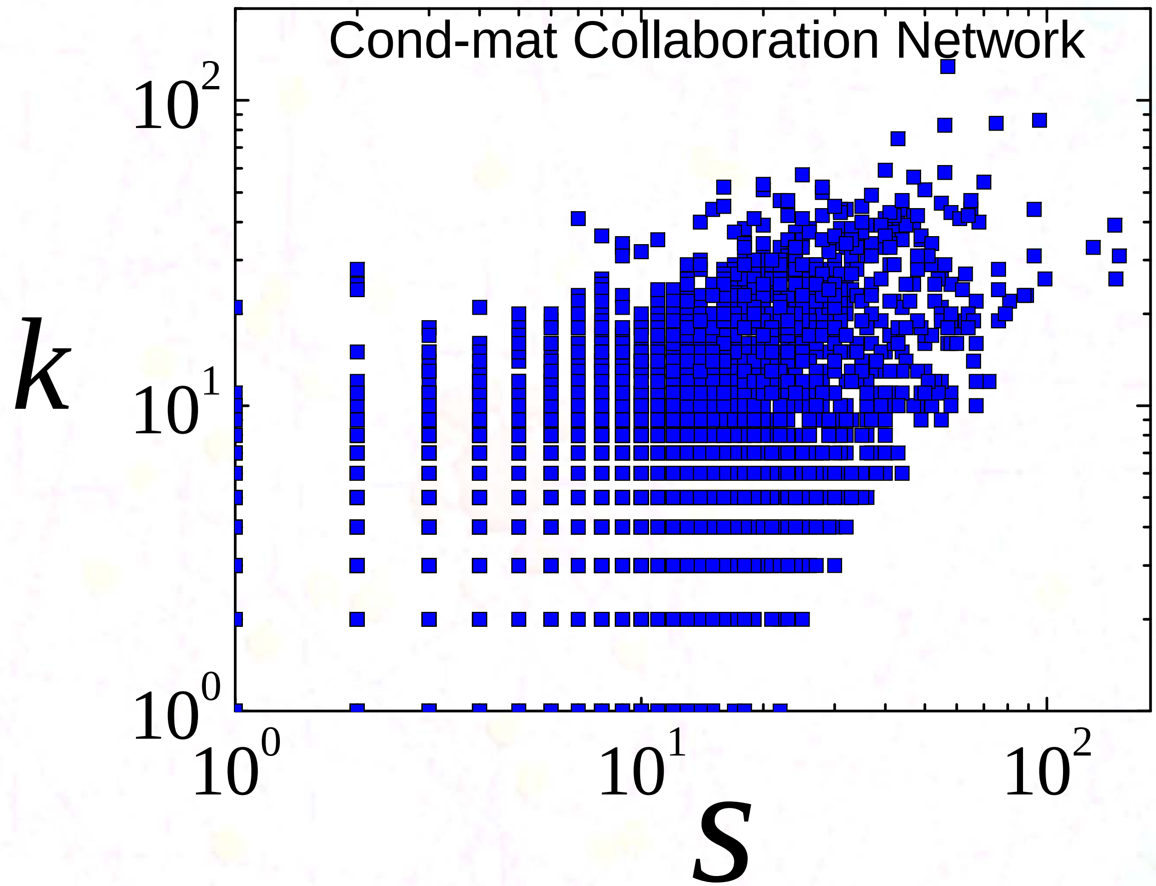
2) Degree in the unipartite projection is proportional to degree in the bipartite network!

$$\bar{k}_{UNI} \sim \kappa \left\langle \lambda^2 \right\rangle \Big/ \left\langle \lambda \right\rangle^2 \qquad P_{UNI}(k) \sim k^{-\gamma}$$

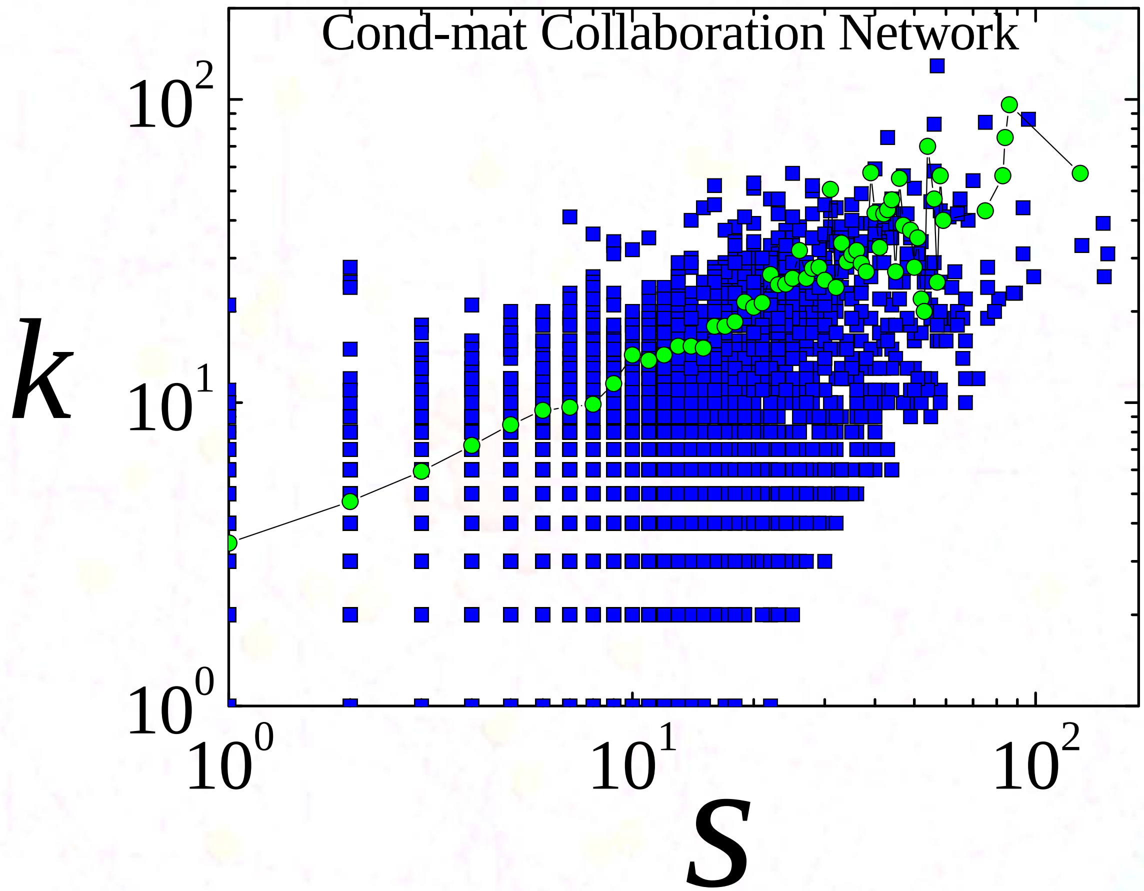


Connectivity of S1S1 is fully controlled by fitness values.

Connectivity of Bipartite Networks (Revisited)

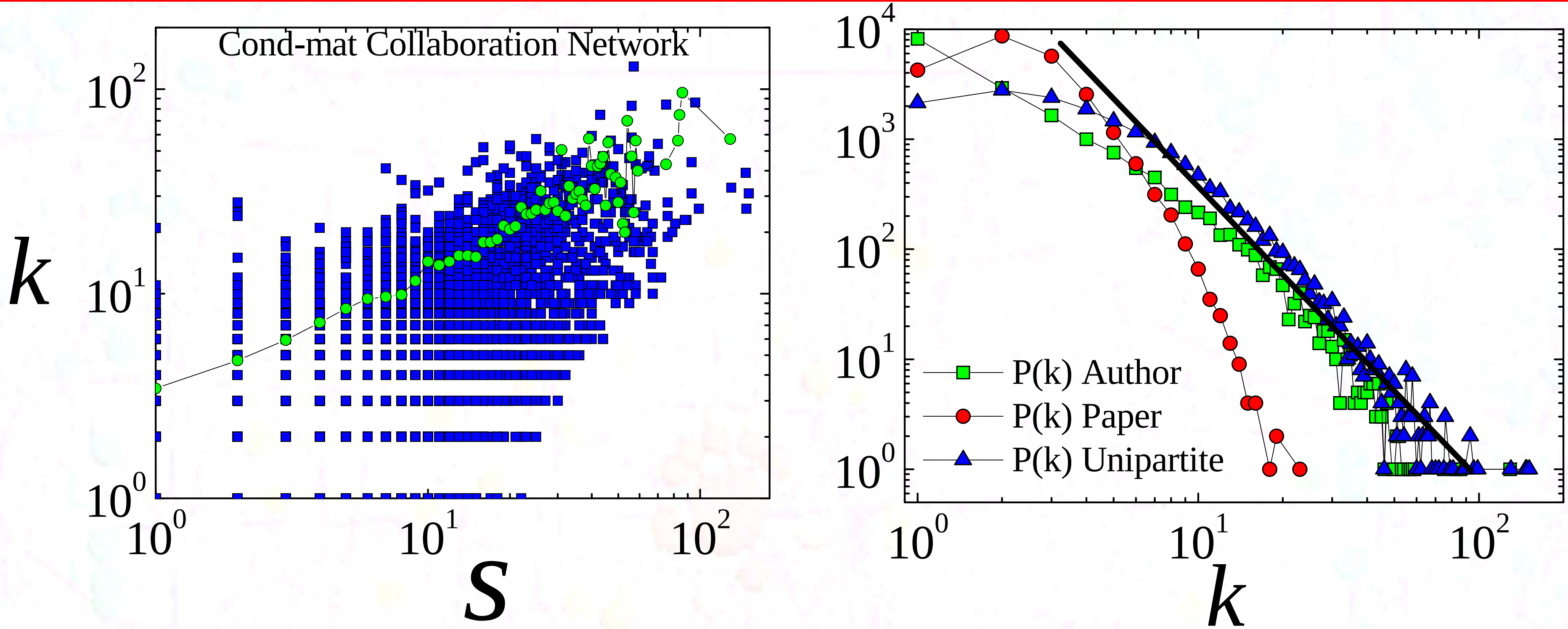


Connectivity of Bipartite Networks (Revisited)



Average projection degree is proportional to author degree

Connectivity of Bipartite Networks (Revisited)

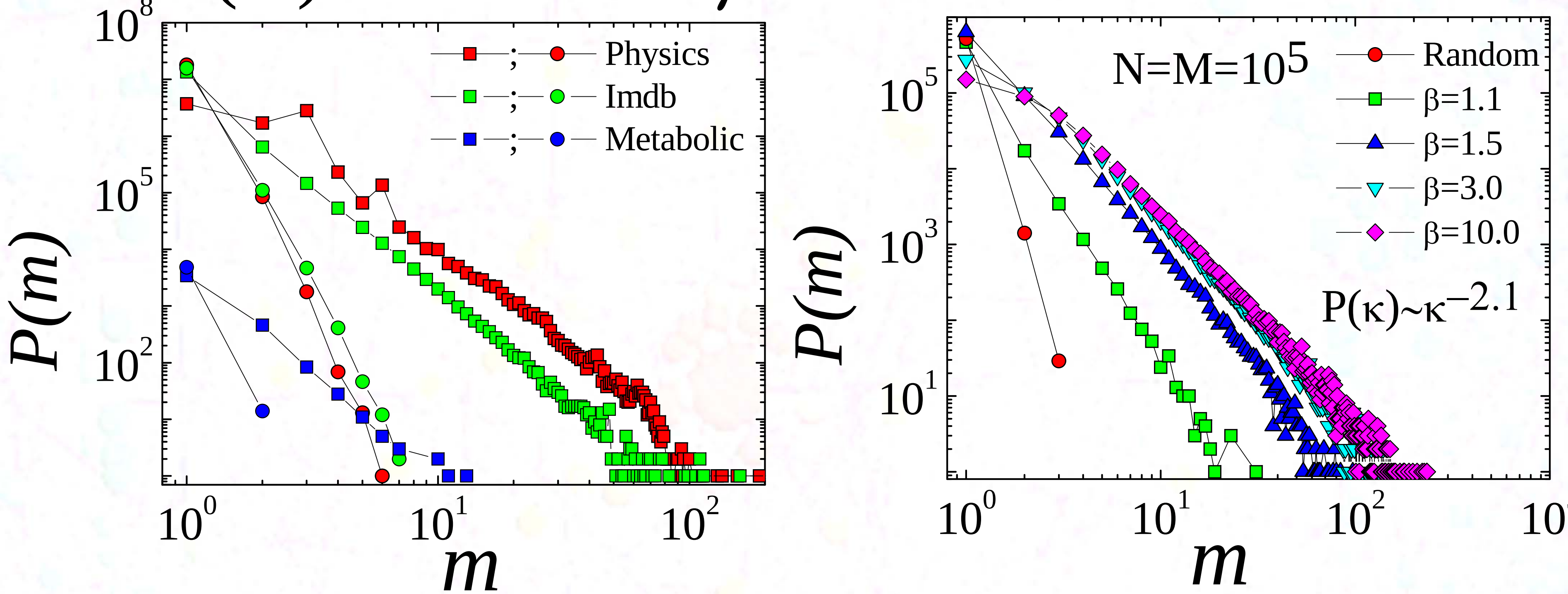


P(k) of the projection is similar to P(k) of authors for large k values.

Scale-free degree distribution of top (bottom) nodes leads to asymptotically scale-free distribution of the corresponding unipartite projection.

How many papers two authors have in common? (Revisited)

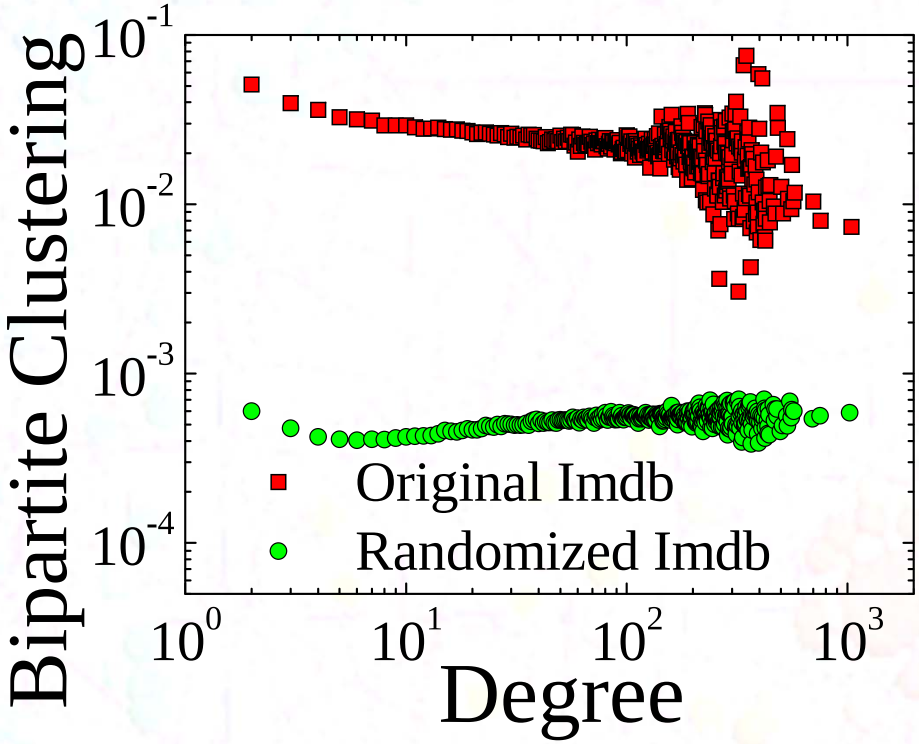
$$P(m) \sim m^{-\tau} \quad \tau = 2\gamma - 3/2 \quad \text{for} \quad \beta = 2$$

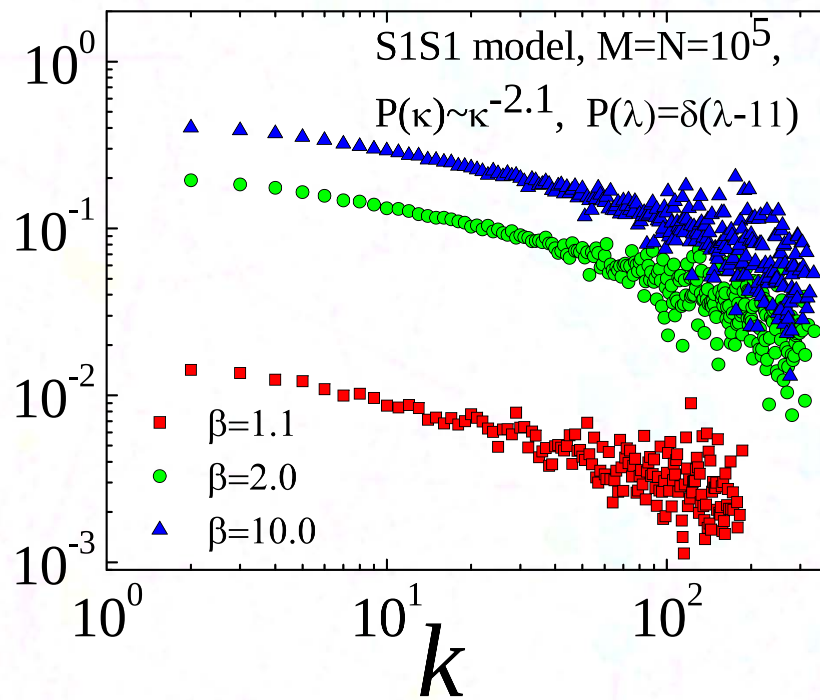


β tunes the power-law exponent!

The fat tail of shared neighbor distribution is the direct consequence of metric property of the space

Bipartite Clustering (Revisited)





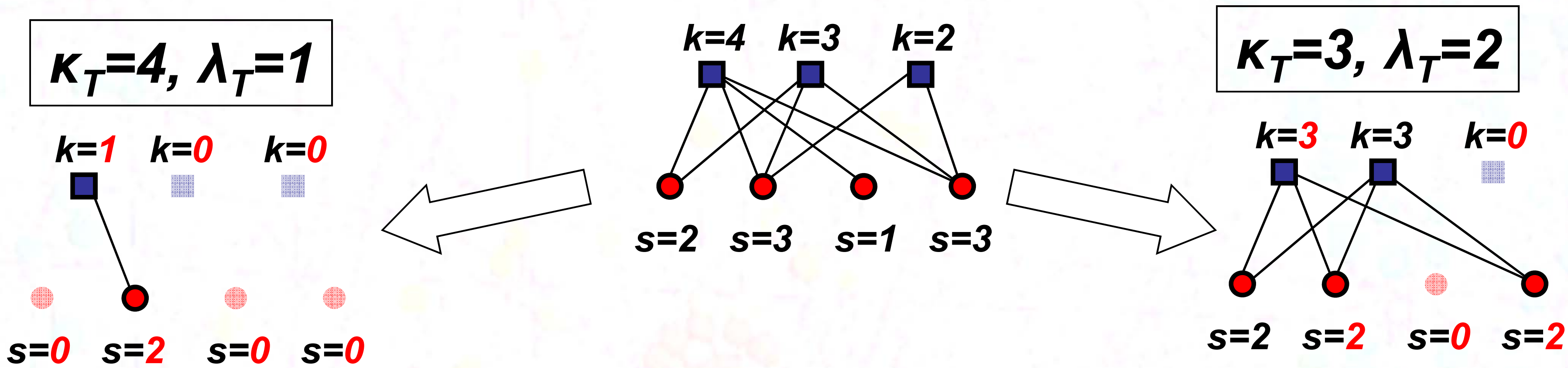
β tunes the bipartite clustering!

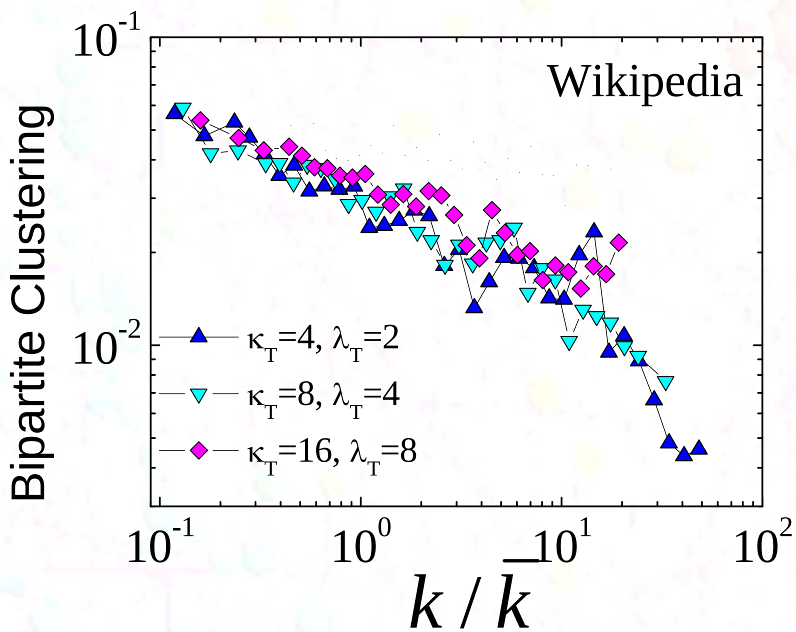
Power-law? **No!** $C_B(k) \sim \frac{Ln(1+ k/k_0)}{k}$ for $\gamma = 3$, $\beta = 2$

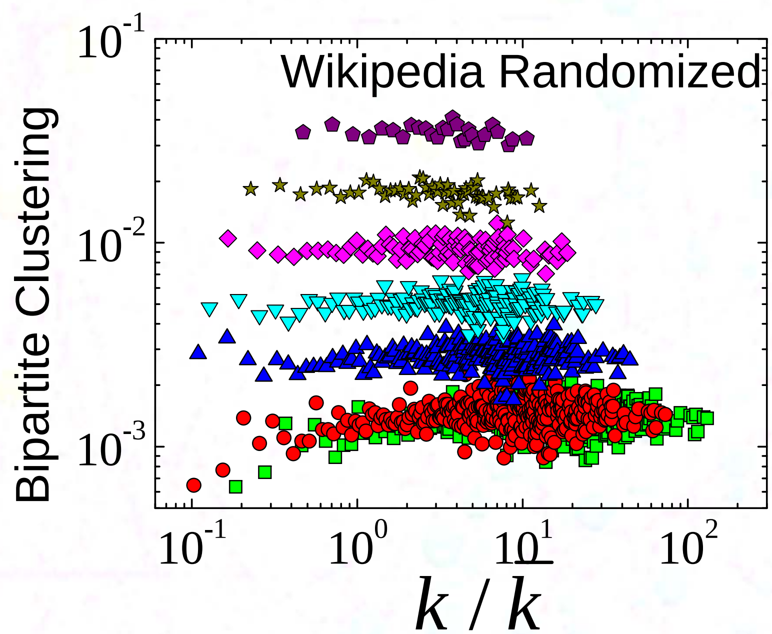
High bipartite clustering is the direct consequence of the metric property of the space

Degree Thresholding "Symmetry"

Remove top and bottom nodes: $k<\kappa_T$; $s<\lambda_T$. Do not Iterate!







Degree Distribution and Clustering are Self-Similar in S1S1

Embedding Bipartite Networks in Metric Spaces

Maximum Likelihood Approach

1) *Determine parameters of the S1S1 model (**β**,**μ**,**R…**)*

2) *Scatter the nodes of the network uniformly over the circle of **R**. Assign fitness values (**κ_i, λ_j**) to nodes according to degree distributions.*

3) *Calculate the likelihood of given network layout to be generated by the S1S1 model:*

$$L = \sum_{i,j} p_{i,j}^{A_{i,j}} \left(1 - p_{i,j}\right)^{1-A_{i,j}}$$

$A_{i,j}$ ***- adjacency matrix***

4) *Start moving nodes on the circle.*
Accept the move only if the likelihood increases.

Summary

1) *High bipartite clustering and power-law distribution of the number of shared neighbors in bipartite networks naturally explained by existence of underlying metric spaces.*

2) *S1S1 models can reproduce most properties of real bipartite networks.*

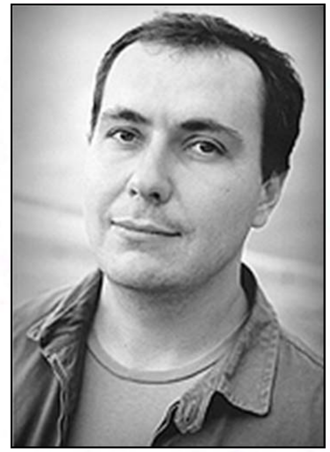
3) *S1S1 models and real bipartite networks are self-similar upon the degree-thresholding renormalization.*

4) ***Challenge:*** *efficient embedding algorithms.*

5) ***Possible Applications:*** *recommendation systems, signalling pathways, content search.*

Collaborators:




F. Papadopoulos
U of Cyprus

M. A. Serrano
U of Barcelona

M. Boguñá
U of Barcelona

D. Krioukov
CAIDA/UCSD

kc claffy
CAIDA/UCSD

Papers:

*M. Boguñá, and R. Pastor-Satorras, PRE **68** 036112 (2003).*
*M. A. Serrano, D. Krioukov, and M. Boguñá, PRL **100** 078701 (2008).*
*M. Boguñá, D. Krioukov and kc claffy, Nature Physics **5** 74 (2009).*
*M. Boguñá and D. Krioukov, PRL **102** 058701 (2009).*
*D. Krioukov, F. Papadopoulos, A. Vahdat, and M. Boguñá, PRE **80** 035101(R) (2009).*

caida
UC San Diego
Local Impact, National Influence, Global Reach
BOSTON UNIVERSITY

Degree Thresholding

Remove top and bottom nodes: $k<\kappa_{T;}$ $s<\lambda_{T.}$ Do not Iterate!