Do Bipartite Networks Have Metric Structure?

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.

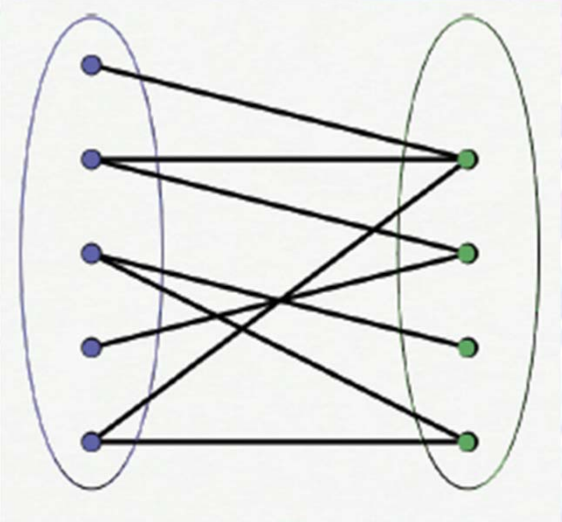
(2) Class of Bipartite Networks in Metric Spaces.

(3) Applications for Bipartite and Non-Bipartite Networks.

(4) Open Questions.

UCSD Complex Network Seminar (DANCES),
La Jolla, May,11, 2011

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.

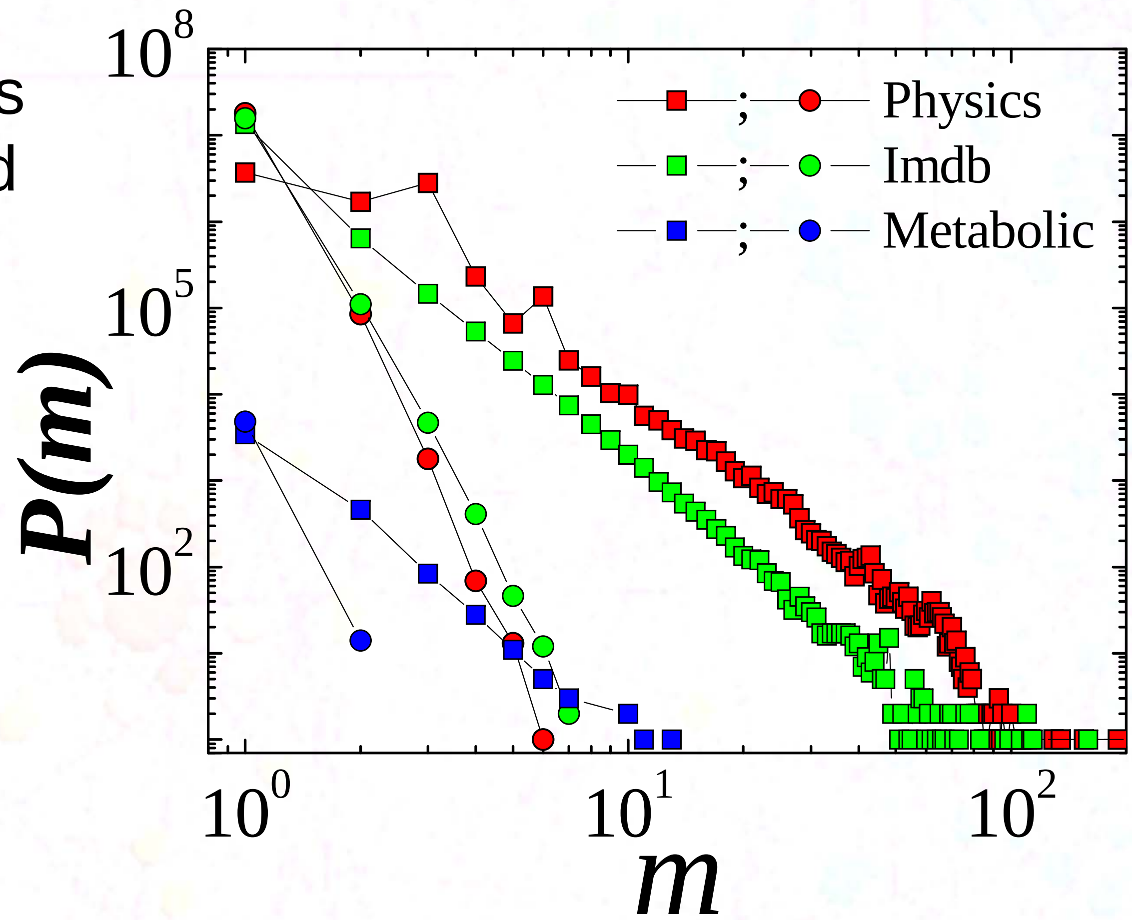
Examples:

- 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)

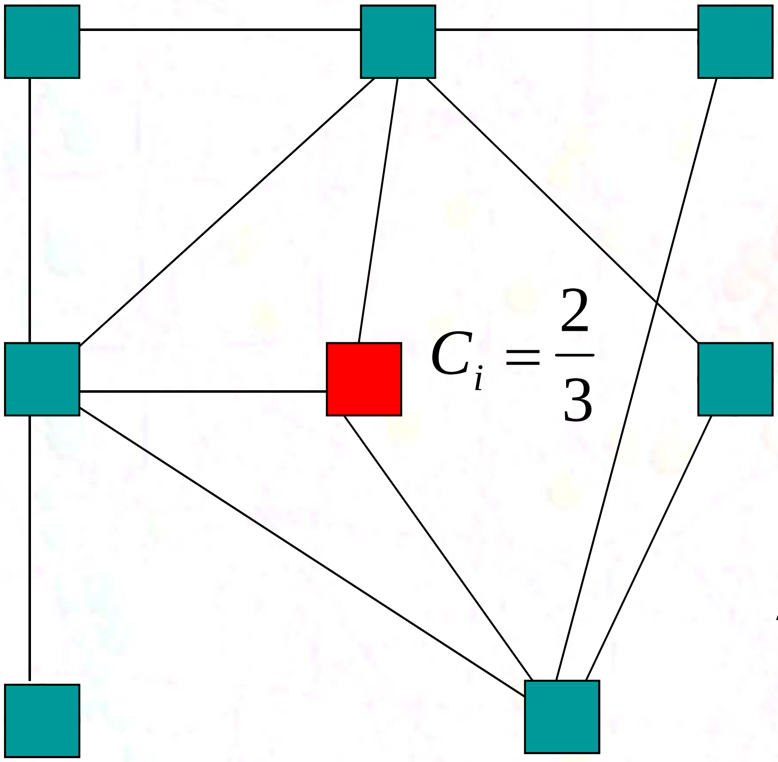


1) P(m) is distributed as a power-law.
2) M is significantly higher in real bipartite networks than in randomized counterparts.

Ordinary Clustering Coefficient

How close are node neighbors from being a complete graph?

$$C_i = \frac{\text{Number of links among neighbors of node } \boldsymbol{i}}{\text{Total possible links among neighbors of node } \boldsymbol{i}}$$



$$C_i = \frac{2}{3}$$

D. J. Watts and S. Strogatz Nature **393** (1998).

$$C_i = \frac{2\sum_{m \neq n} e_{mn}}{k_i\left(k_i - 1\right)}$$

Average clustering for degree *k*:

$$\overline{C}(k) = \langle C_i \rangle_{k_i = k}$$

Average clustering for the network:

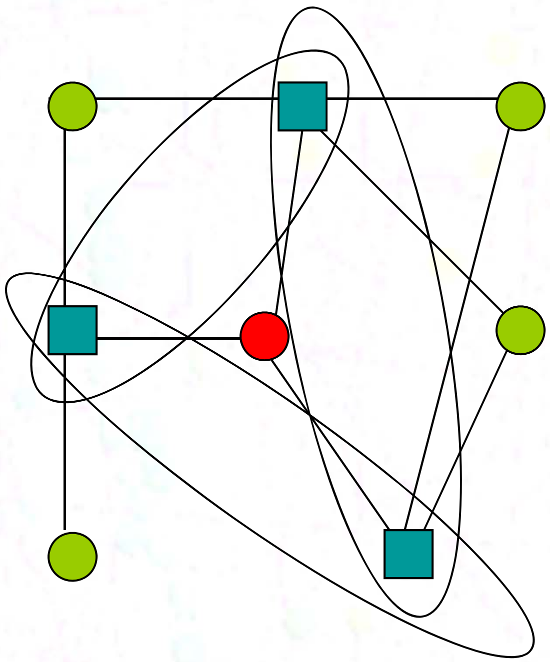
$$\overline{C} = \frac{1}{N}\sum_i C_i$$

Many real networks are highly clustered!
(compared to their random counterparts)

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!

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



Consider all pairs of neighbors

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

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

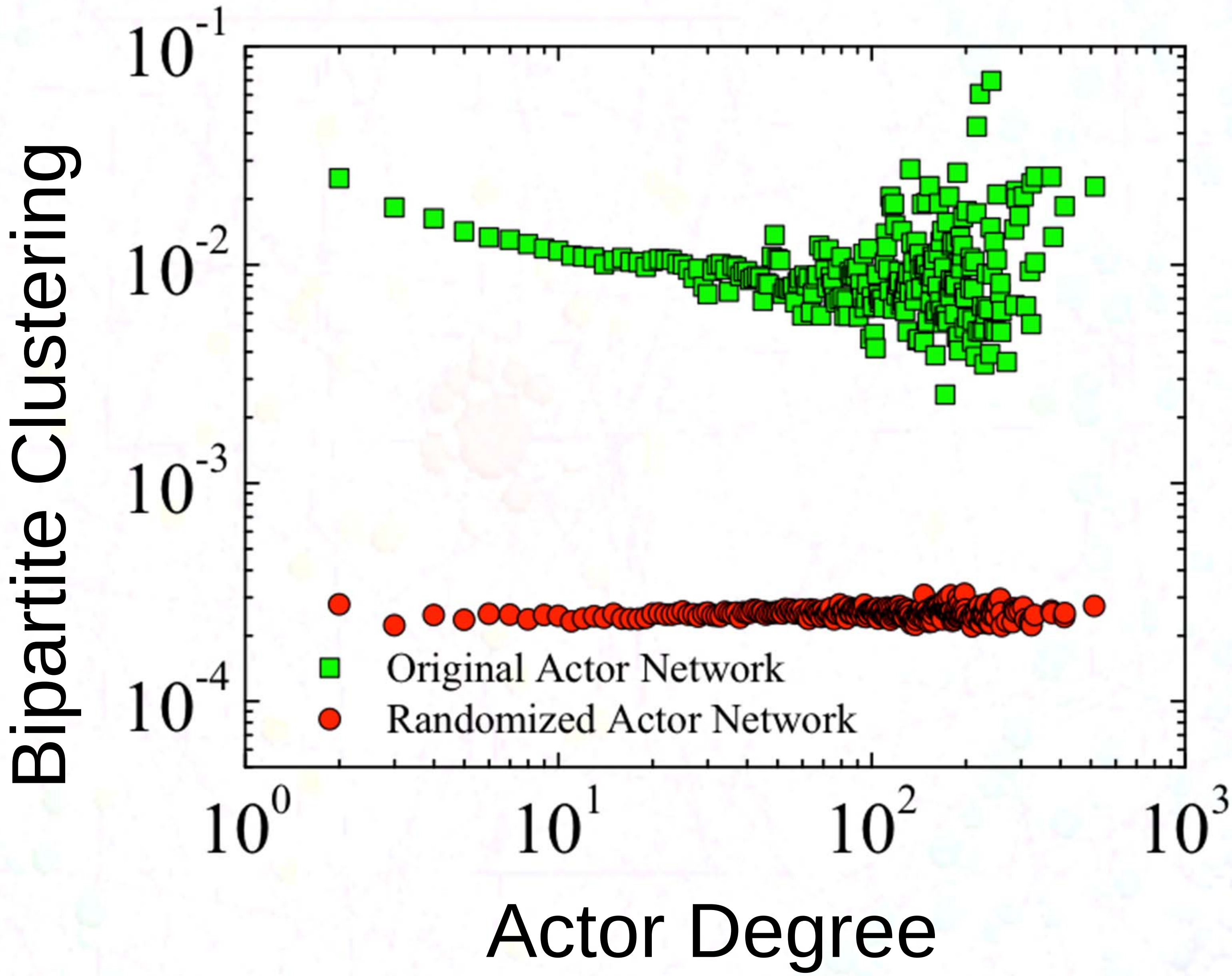
2 Nodes in common, **3** nodes altogether.

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

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

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

Real bipartite networks are highly clustered



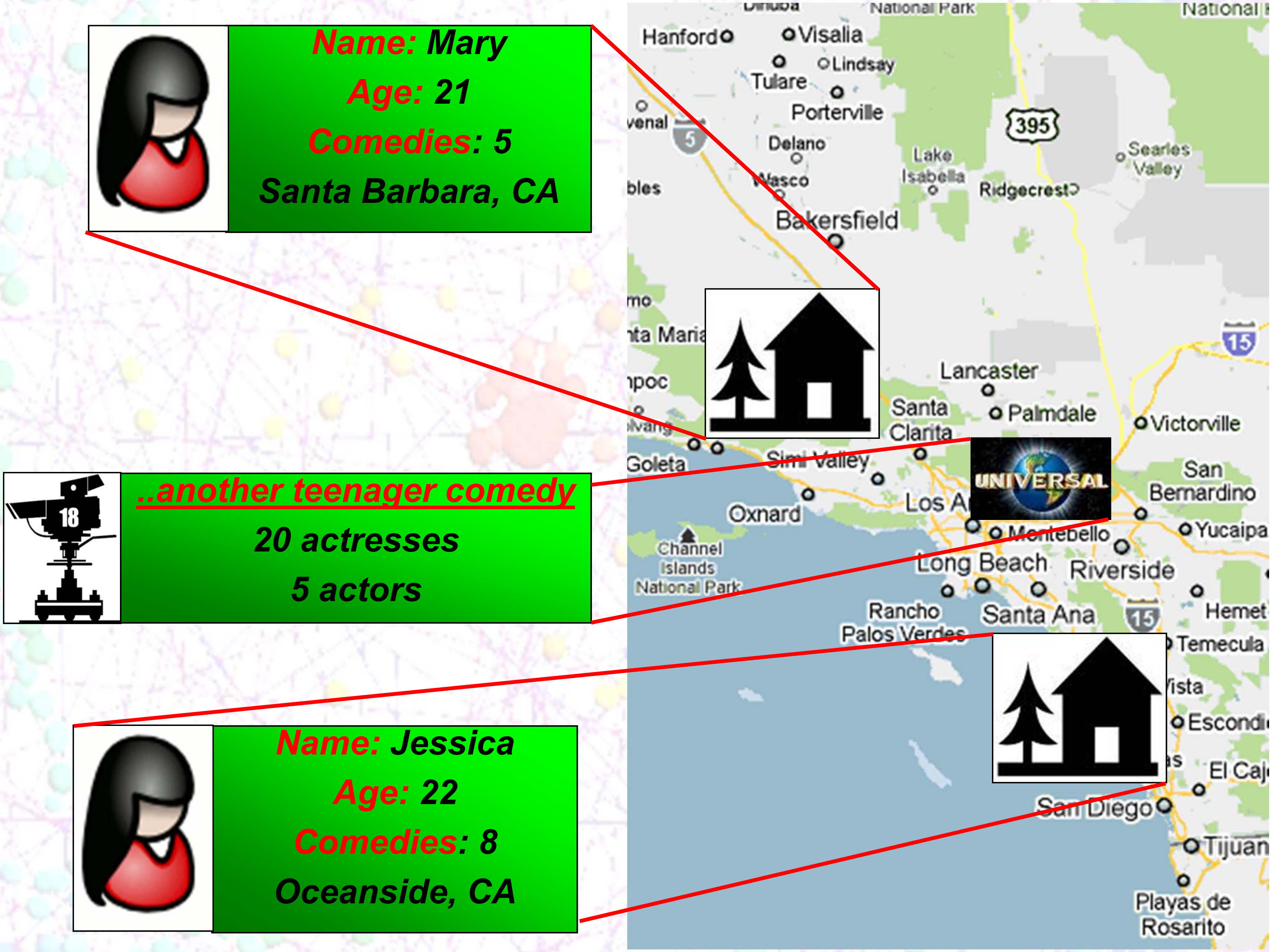
Properties of Bipartite Networks

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?



Actor network has metric structure?

A. *smaller distances imply higher connection probability!*

B. *small* d_1 *and* d_2 *imply small* d_3*?*

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

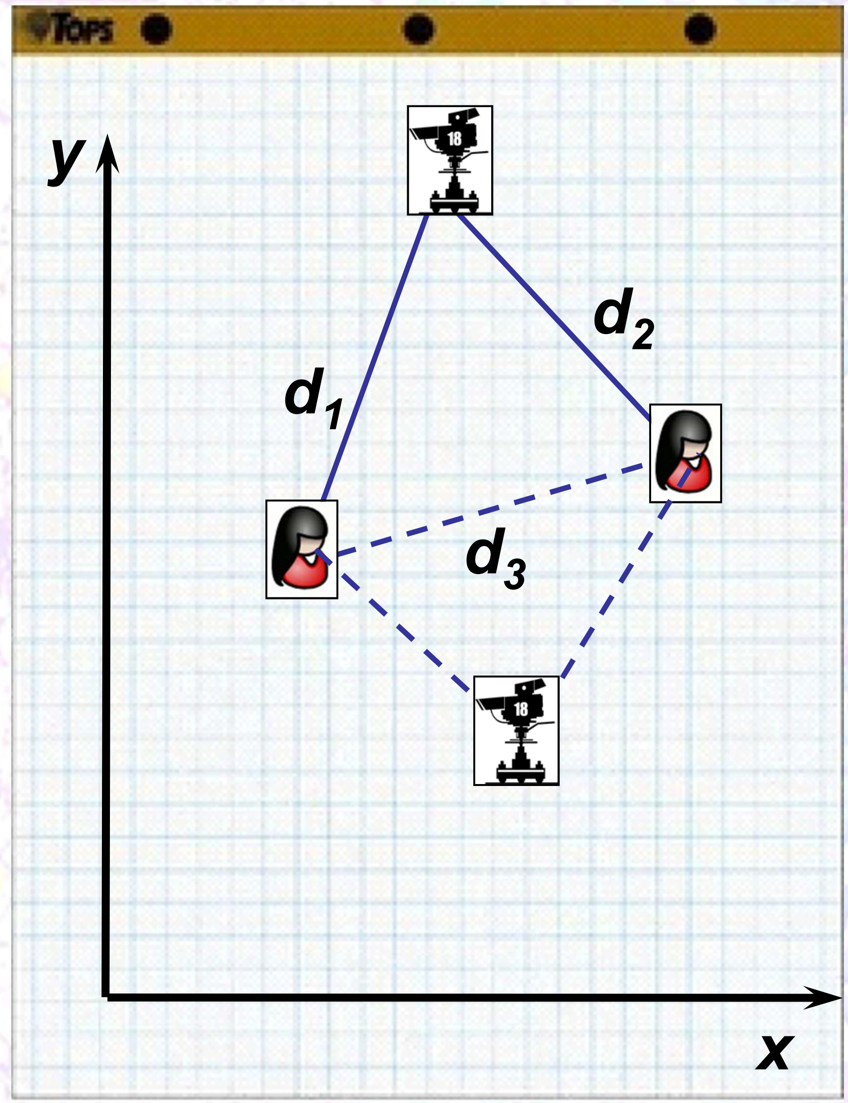
Underlying Space is Metric!

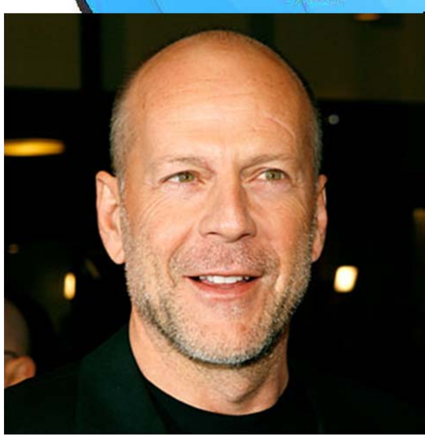
1) $d(x,y) \geq 0$

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: ***56***

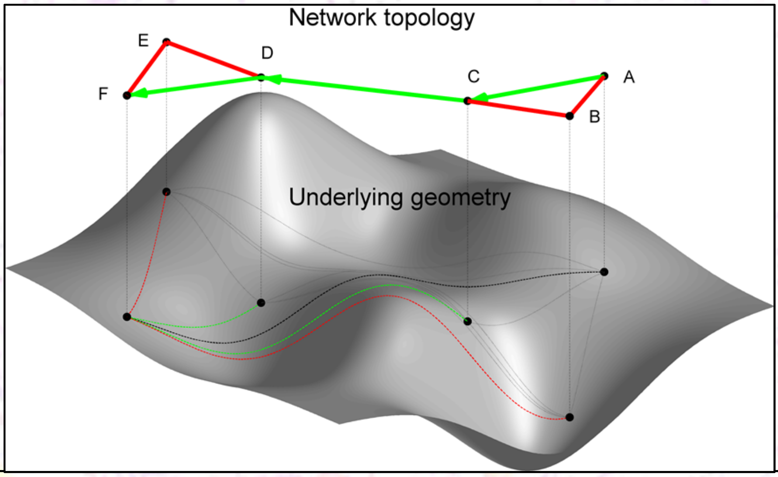
Movies: 82 (as of 2010)

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.* $(d_3 \leq d_1 + d_2)$

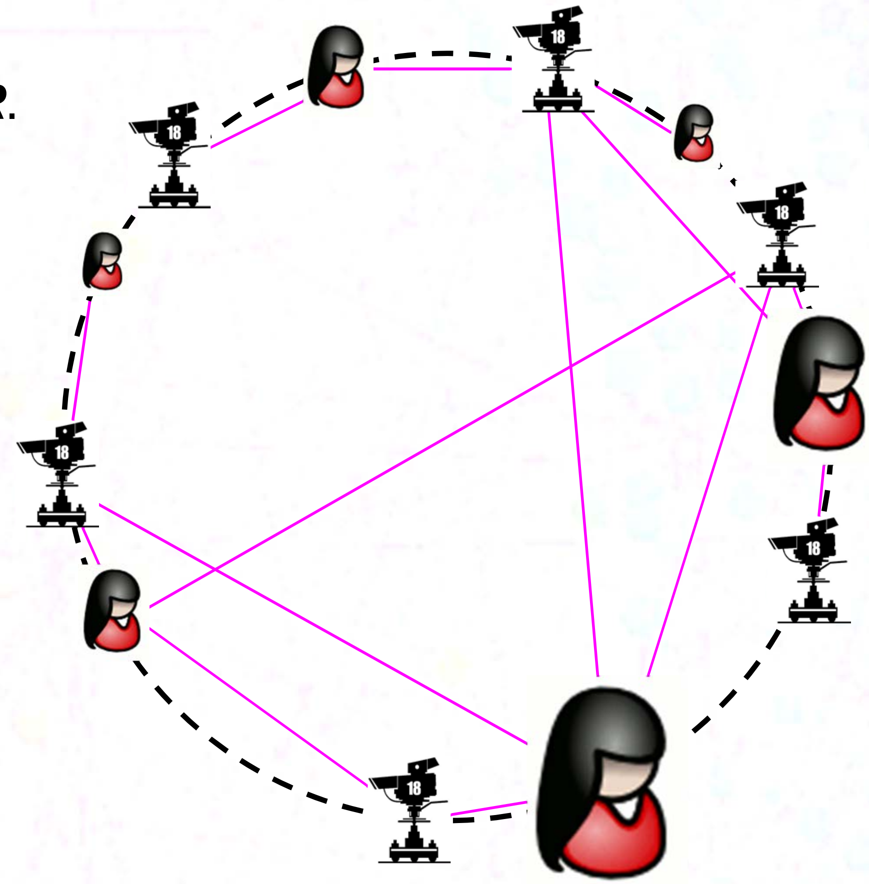
B. *The probability of a link connecting a pair of nodes determined by distance between the nodes in this space.*

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 hidden variable:*
κ *(top nodes)* λ *(bottom nodes). Distance scale:* $d_c = d_c(\kappa\lambda)$

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 node of each kind calculate its hidden variable drawn from ***P(κ) or P(λ)***.

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

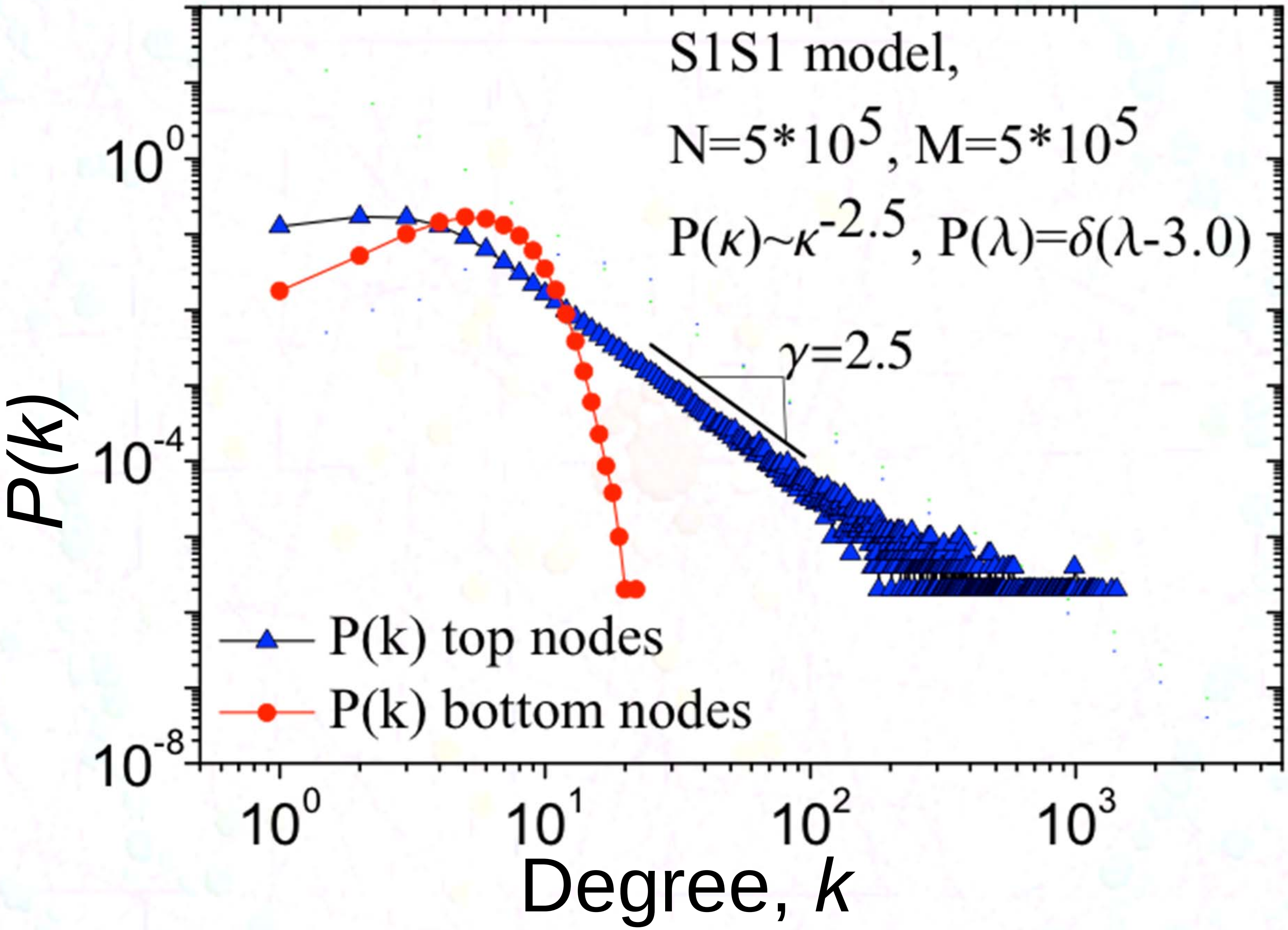
$$P(\lambda) \sim \lambda^{-\gamma_2}; P(\lambda) = \delta(\lambda - \lambda_0)$$

3) Connect authors and papers with probability

$$p_{i,j} = \frac{1}{1 + {x_{ij}}^{\beta}} \qquad x_{i,j} = \frac{d_{ij}}{d_c(\kappa_i \lambda_j)}$$

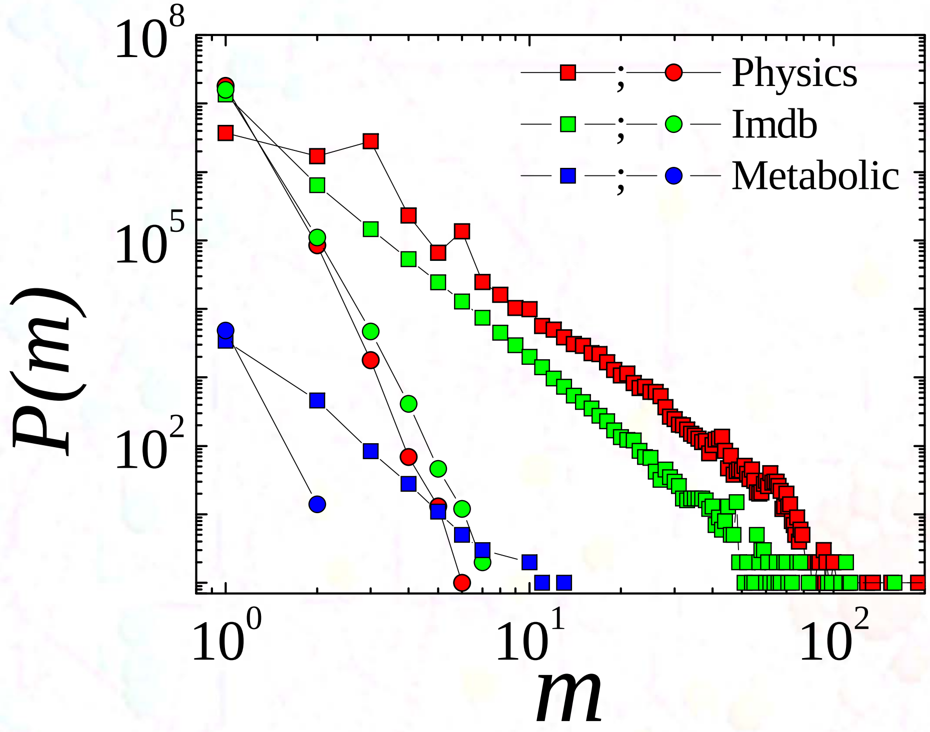
Large ***β*** *corresponds to preferred connections at small distance*

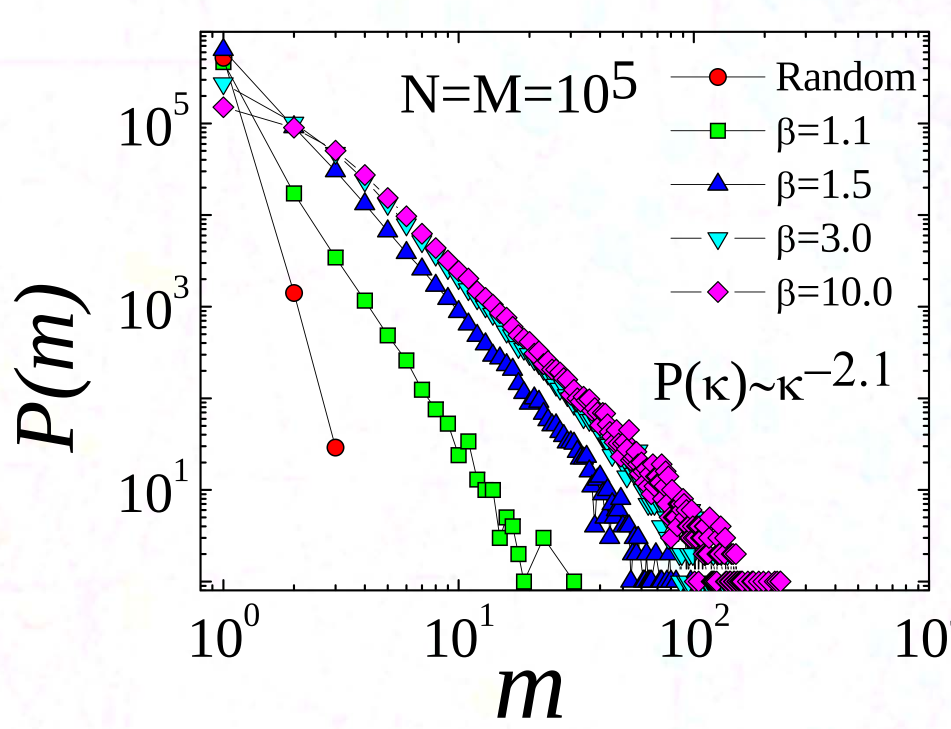
Connectivity of S1S1 model.



Connectivity of S1S1 is fully controlled by hidden variables

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





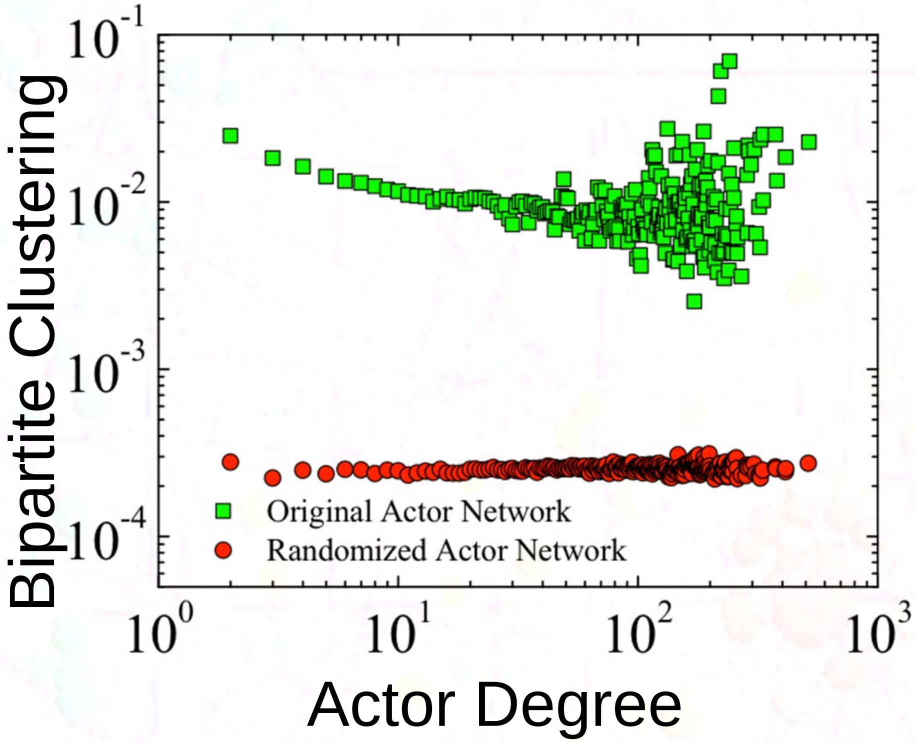
$$p_{i,j} = \frac{1}{1 + x_{ij}^{\beta}} \quad x_{i,j} = \frac{d_{ij}}{d_c(\kappa_i \lambda_j)}$$

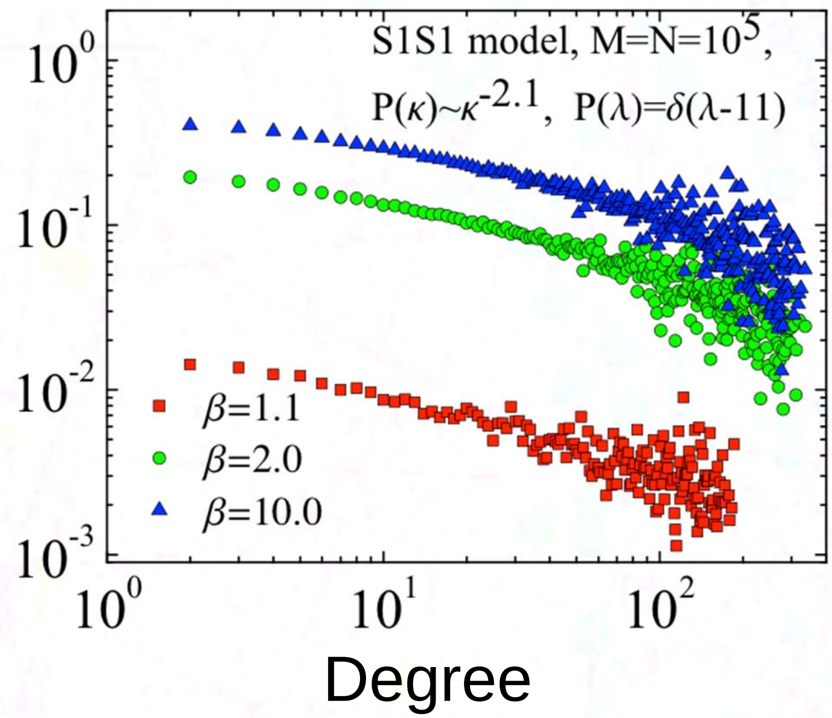
$$P(m) \sim m^{-\tau}$$

$$\tau = \tau(\beta, \gamma)$$

Power law statistics of common neighbors is the consequence of metric property of the space

Bipartite Clustering (Revisited)





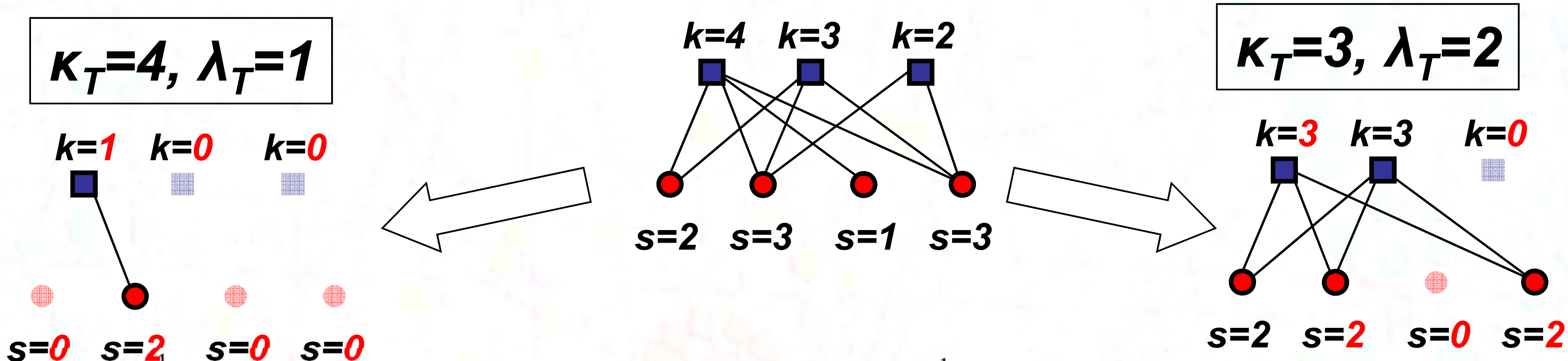
$$p_{i,j} = \frac{1}{1 + x_{ij}^{\beta}} \quad x_{i,j} = \frac{d_{ij}}{d_c(\kappa_i \lambda_j)}$$

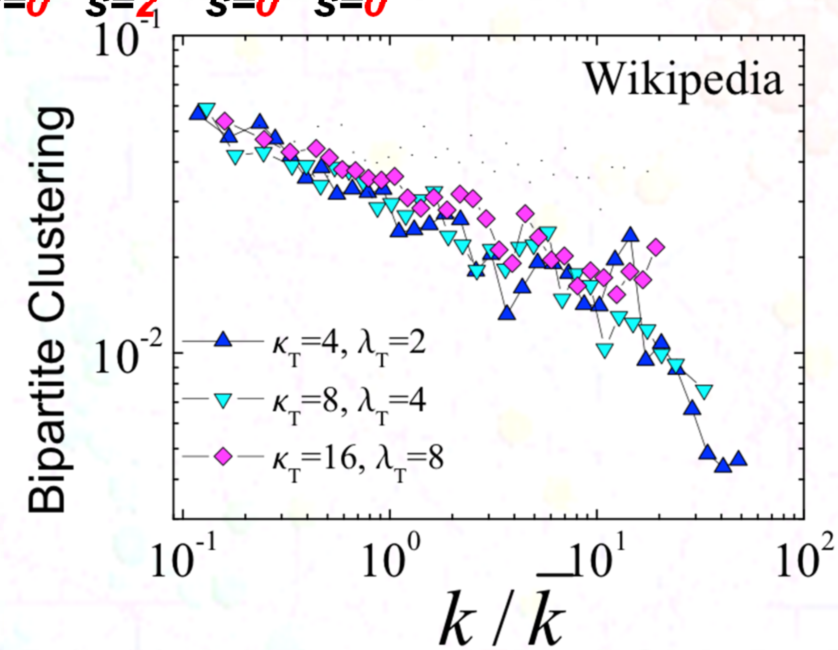
β tunes the bipartite clustering!

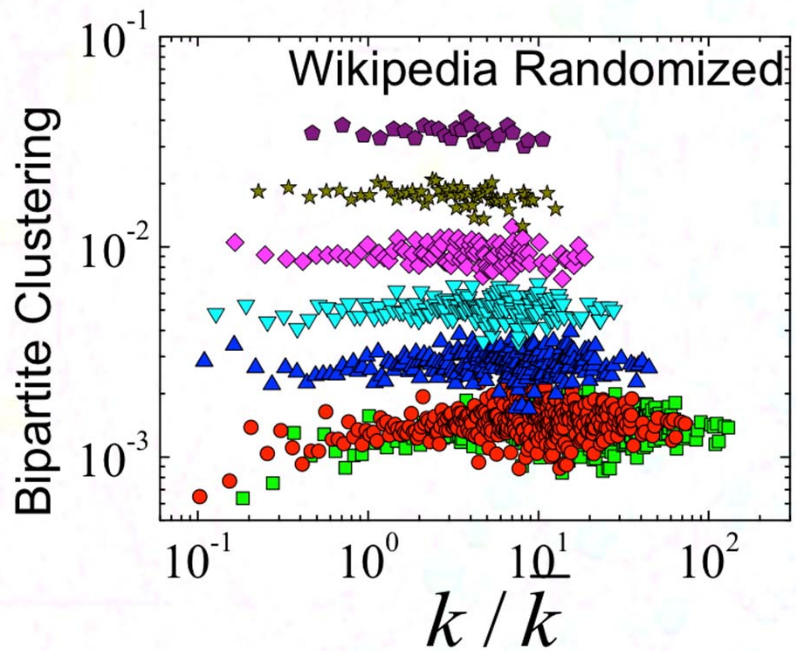
High bipartite clustering
is the direct consequence of the metric property of the space

Degree Thresholding "Symmetry"

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

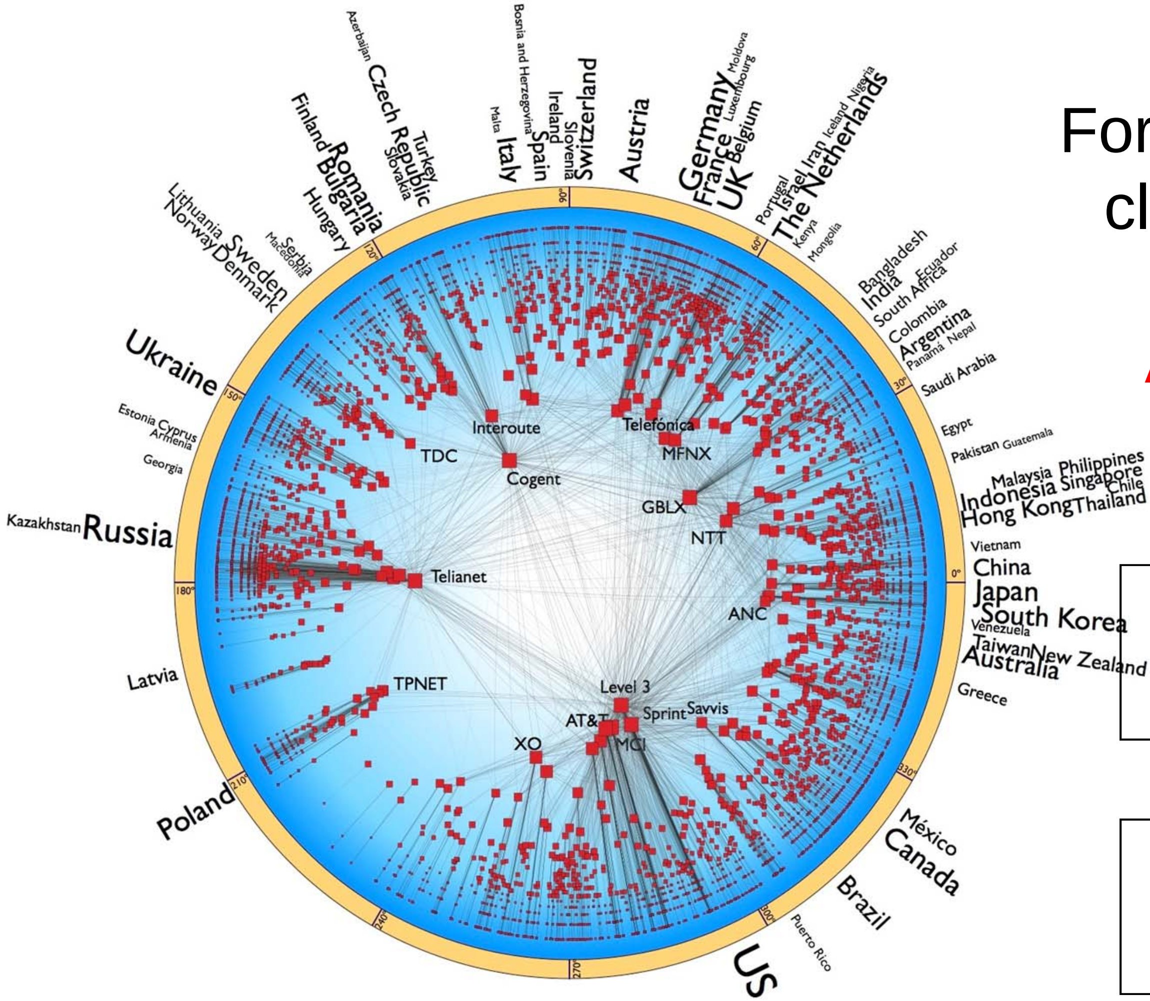






Degree Distribution and Clustering are Self-Similar in S1S1

Applications: AS Internet (Greedy Routing)



Greedy routing:
Forward packets to neighbor closest to the destination.
(in the metric space)
ASes need only local information!

Success Ratio: 97%
Average Stretch: 1.1

30% of failed Ases:
Success Ratio: 85%

M. Boguñá, F. Papadopoulos, and D. Krioukov,
Sustaining the Internet with Hyperbolic Mapping,
Nature Communications, v.1, 62, 2010

Applications for Bipartite Networks

A. **Recommendation Systems:**
Consumers are connected to products they purchased.
Can we recommend new products to consumers?

B. **Metabolic Networks**
Metabolites are related to chemical reactions.
Can we predict missing reactions?
M.Boguna, M.A. Serrano (in preparation) (2011).

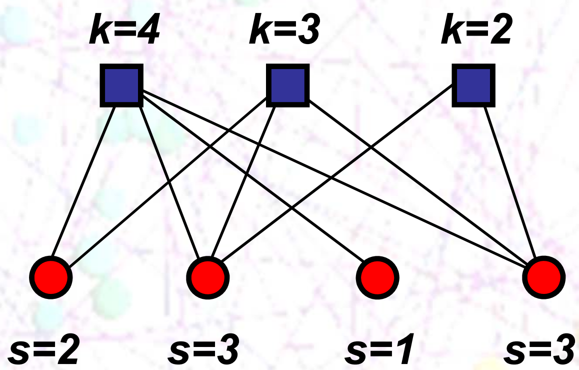
C. **Gene regulatory networks.**

D. **Protein Protein Interaction networks.**

E. **??? Feel free to suggest!**

Open Questions

Why nodes of both kinds belong to the same metric space?



Links between nodes of different kinds.

↓

Only need distances between different kinds of nodes?

Suppose, both sets of nodes are in the same metric space.
Only distances between different kinds of nodes known.
Necessary/Sufficient conditions to infer remaining distances?

Efficient Algorithms to infer coordinates of nodes?
Currently available algorithm is based on maximum likelihood techniques $O(N^3)$. Approximate embedding $O(N^2)$.

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.*

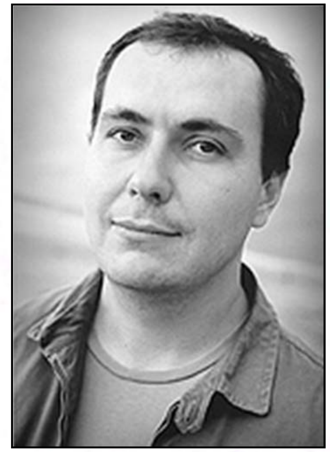
3) *S1S1 models and real bipartite networks are self-similar upon the degree-thresholding renormalization.*

4) ***Challenge:*** *efficient embedding algorithms. Currently available algorithm $O(N^3)$. Approximate algorithm $O(N^2)$.*

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).*
D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M.Boguñá, PRE ***82*** *036106 (2010).*
M. Kitsak and D. Krioukov, arXiv:1104.3184

caida

UC San Diego
Local Impact, National Influence, Global Reach

BOSTON UNIVERSITY