



Modeling folksonomy

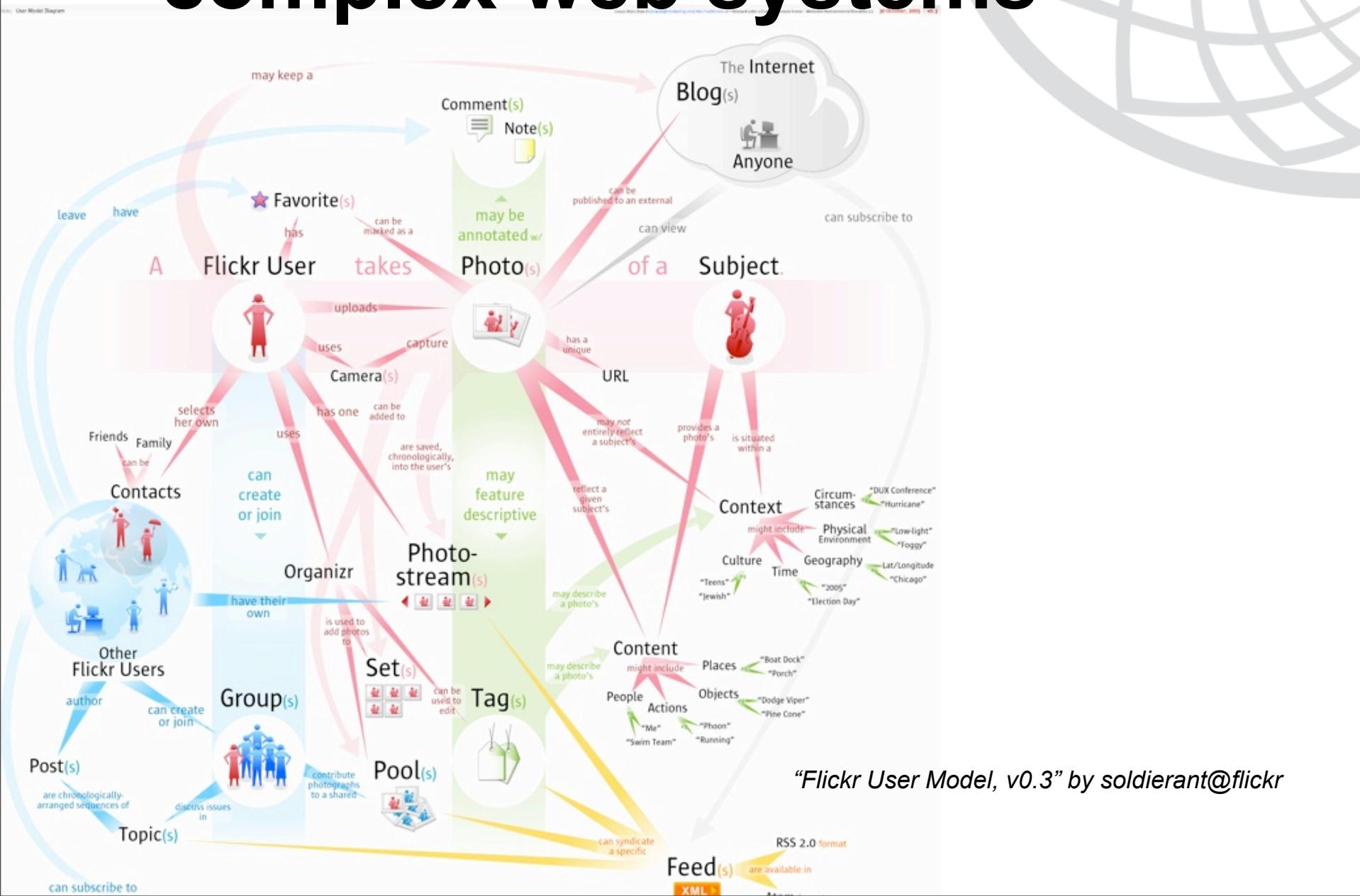
C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, V. Loreto



Hypertext 09, Turin 29 June 2009



Tagora complex web systems

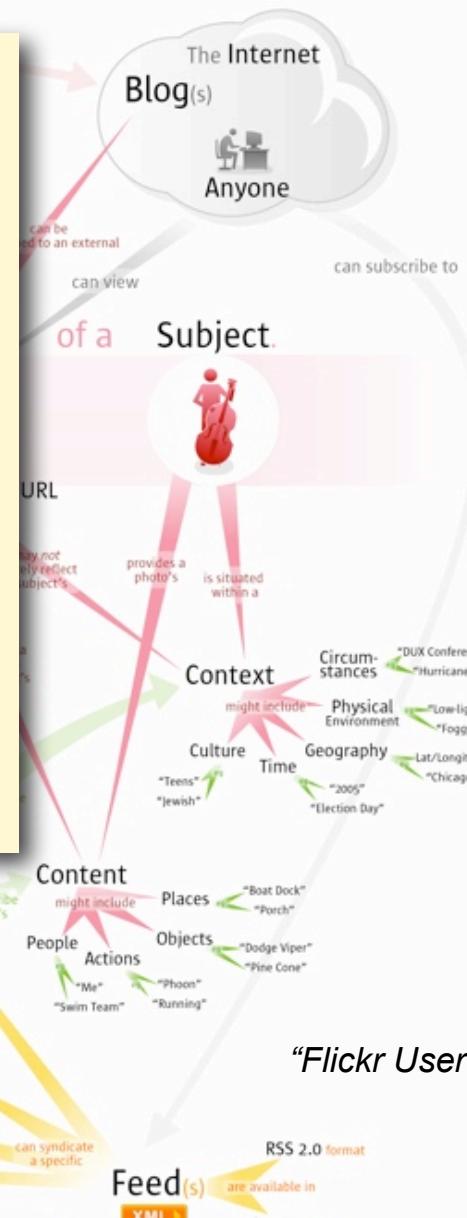
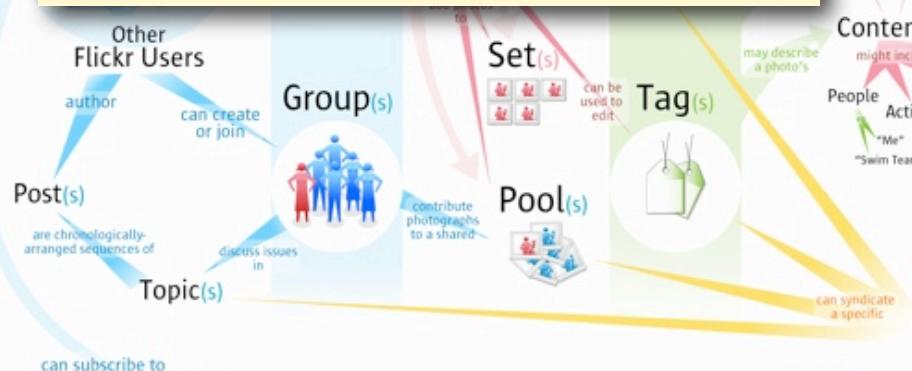


"Flickr User Model, v0.3" by soldierant@flickr

Tagora complex web systems

how?

*more information
less spam
smart interface
more users*

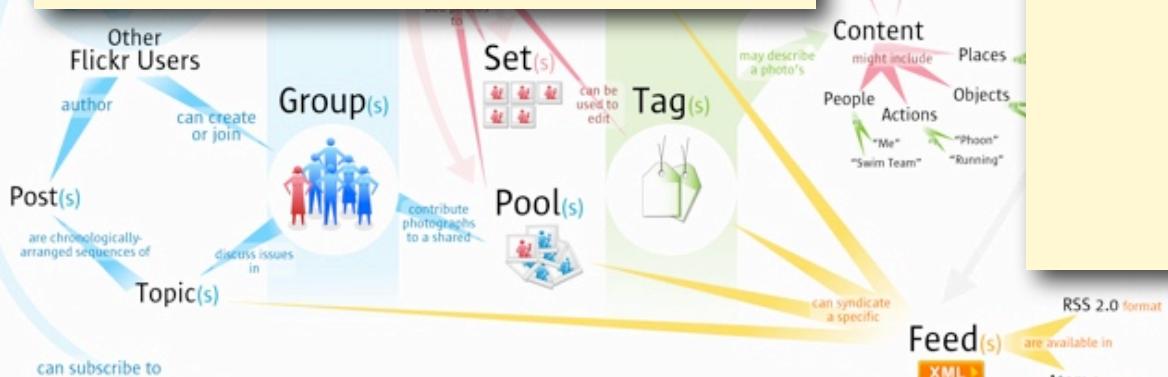


"Flickr User Model, v0.3" by soldierant@flickr

Tagora complex web systems

how?

*more information
less spam
smart interface
more users*



physicist approach:

why?

*statistics
regularities
emergence
universality*

folksonomy tripartite structure



folksonomy tripartite structure



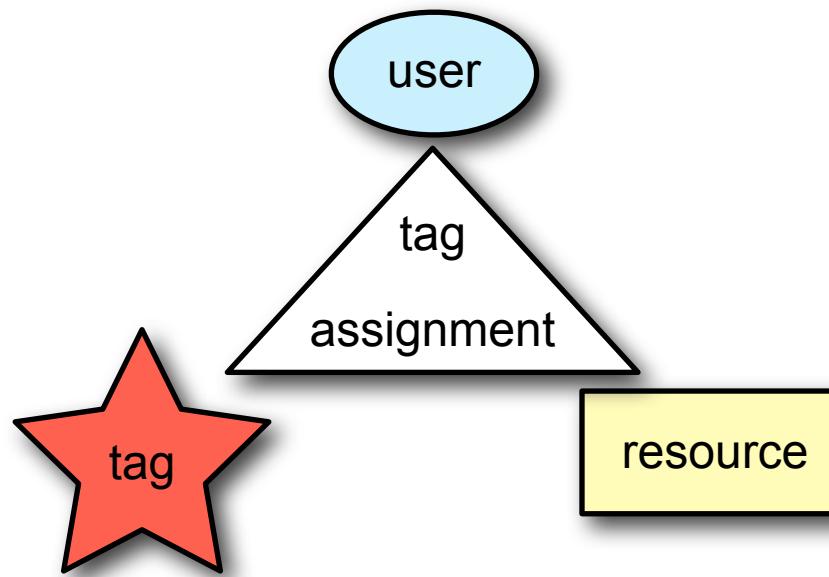
folksonomy tripartite structure

user

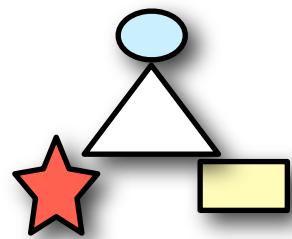
tag

resource

folksonomy tripartite structure

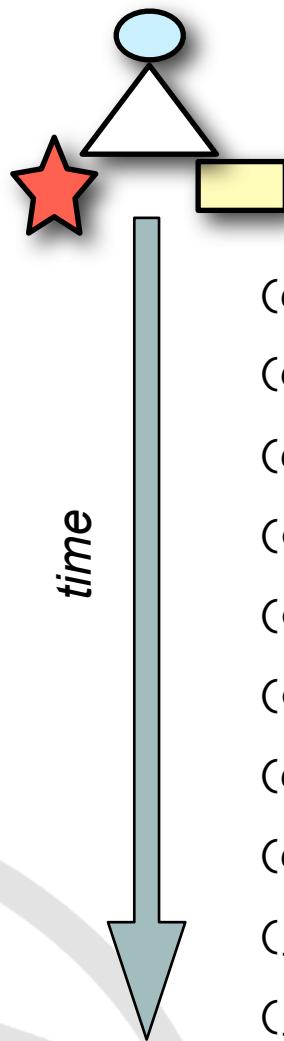


stream view



(user, resource, tag)

stream view



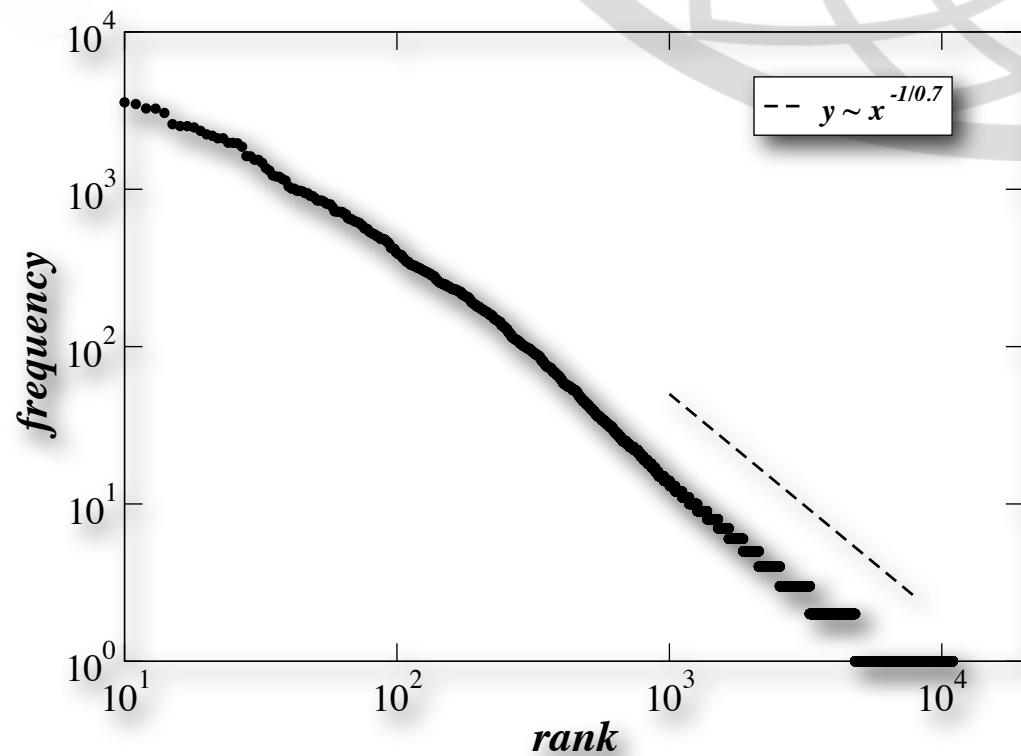
(user, resource, tag)

- (andreab, <http://del.icio.us>, bookmarking)
- (andreab, <http://del.icio.us>, sharing)
- (andreab, <http://del.icio.us>, collaborative)
- (ccattuto, <http://del.icio.us>, folksonomy)
- (ccattuto, <http://del.icio.us>, useful)
- (ccattuto, <http://del.icio.us>, url)
- (andreab, <http://flickr.com>, folksonomy)
- (andreab, <http://flickr.com>, photos)
- (john, <http://tagora-project.eu>, research)
- (john, <http://tagora-project.eu>, folksonomy)
- (john, <http://tagora-project.eu>, eu)

Tagora stream features

Data from del.icio.us (very similar results for bibsonomy.org)

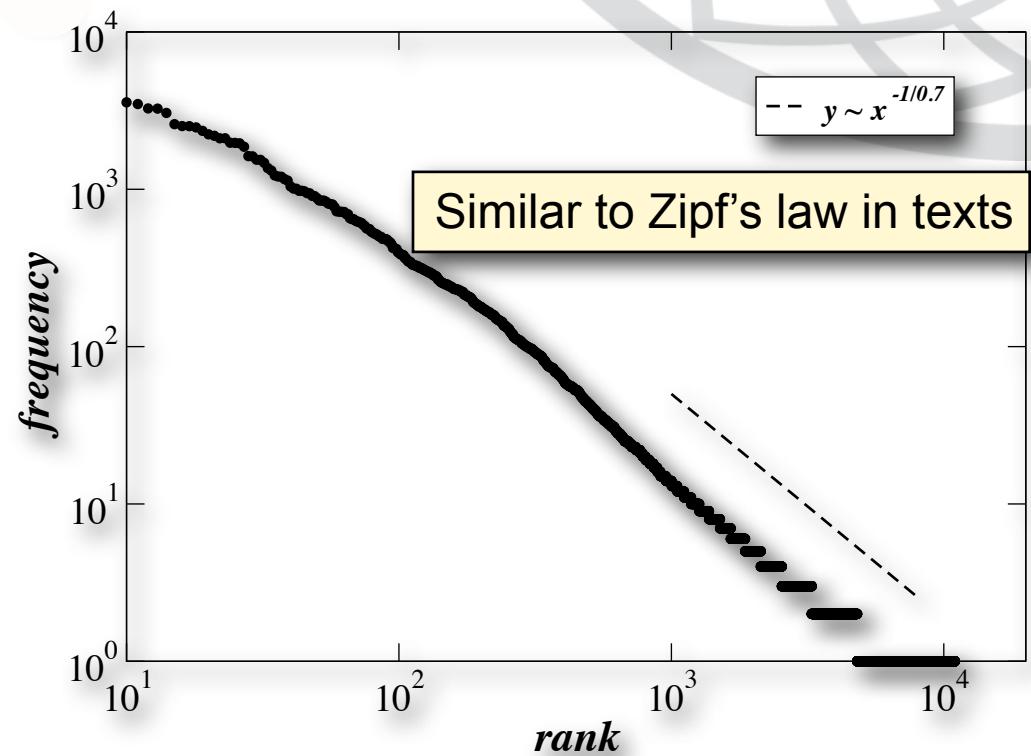
- tag frequencies



Tagora stream features

Data from del.icio.us (very similar results for bibsonomy.org)

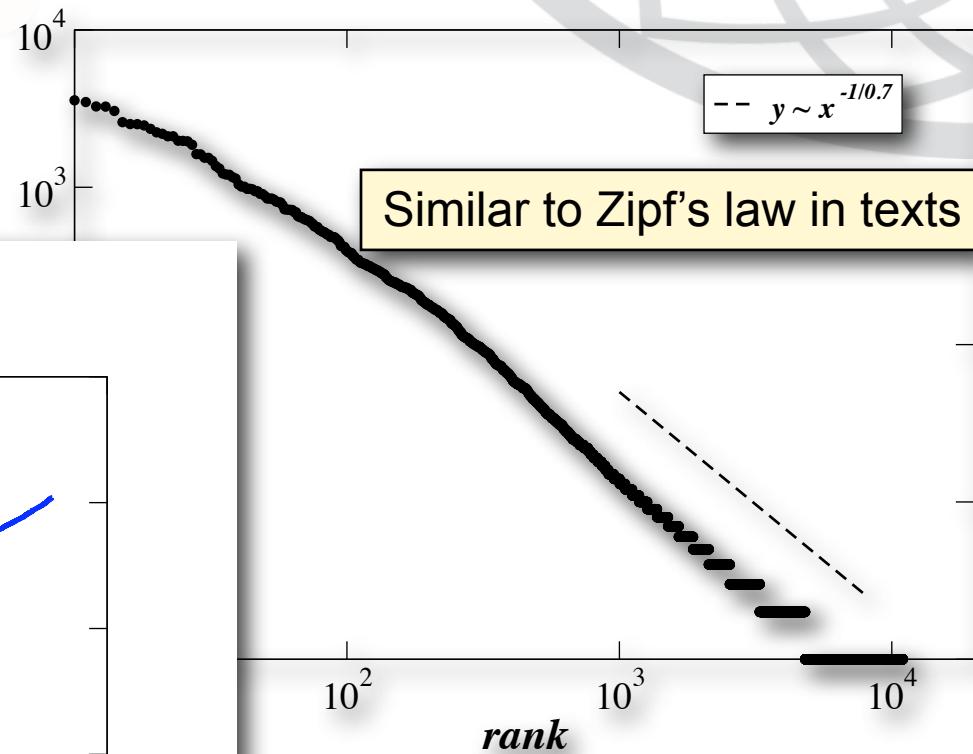
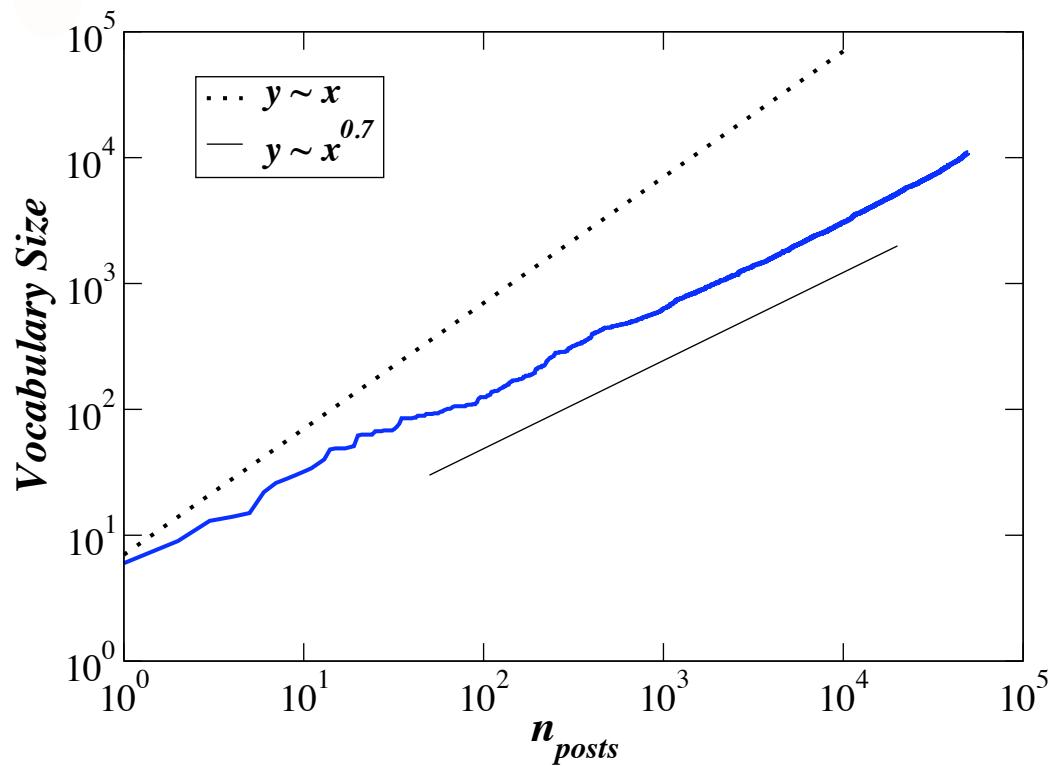
- tag frequencies



Tagora stream features

Data from del.icio.us (very similar results for bibsonomy.org)

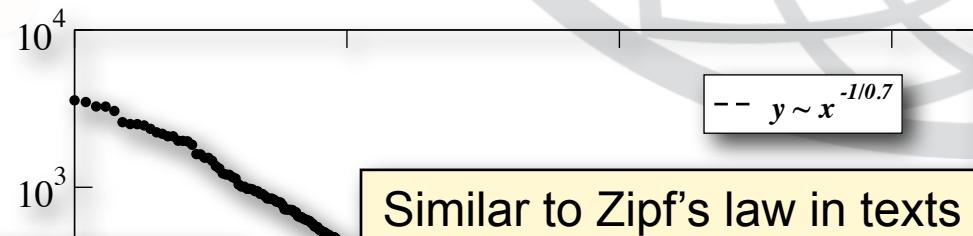
- tag frequencies
- dictionary size



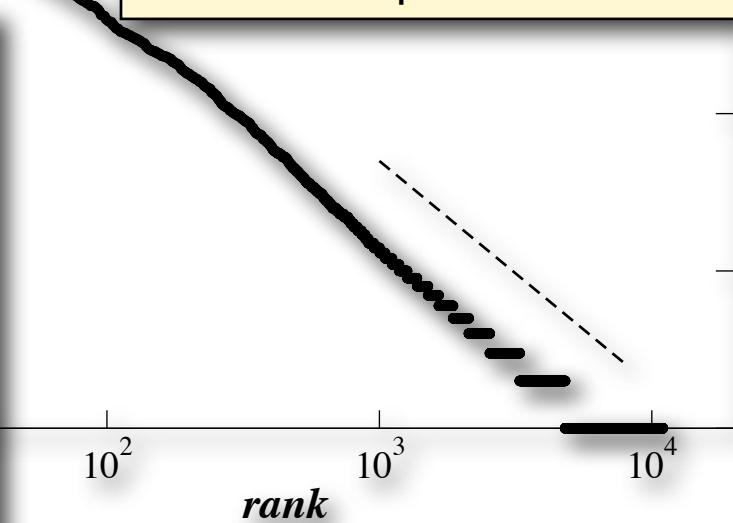
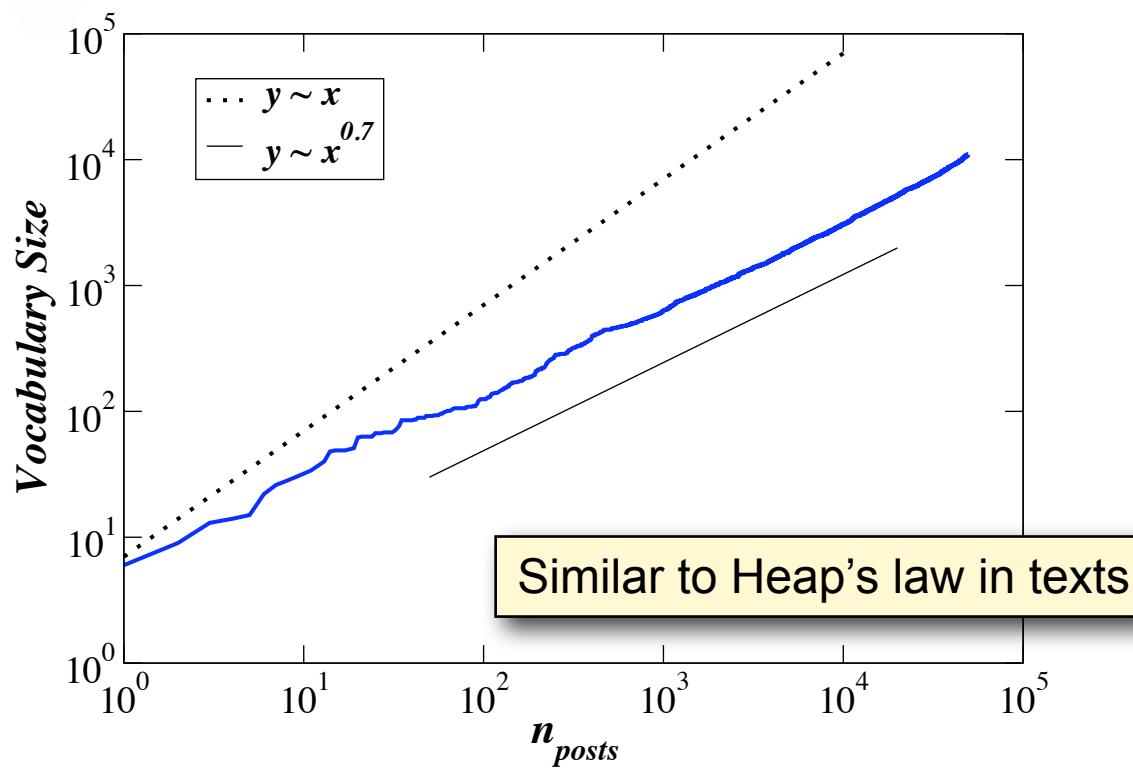
Tagora stream features

Data from del.icio.us (very similar results for bibsonomy.org)

- tag frequencies
- dictionary size



Similar to Zipf's law in texts



Information Society
Technologies

Tagora

tags & semantics?

- hierarchical organization of tags?
- similarity measures?

evaluation with wordnet

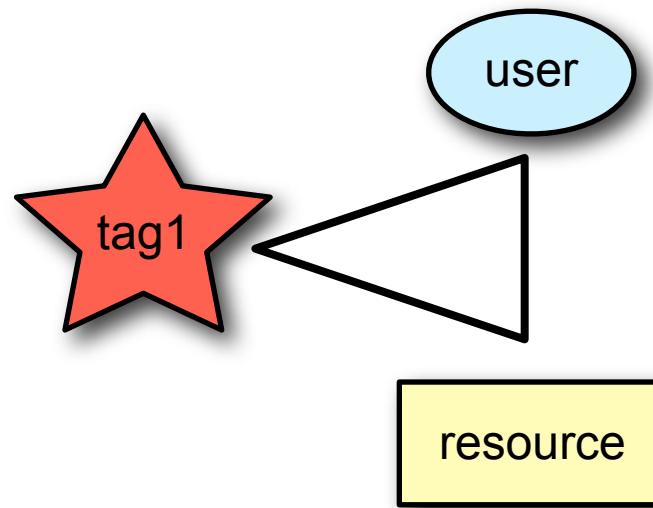
Heynman & Garcia-Molina

Category	freq	folkrank	cosine
1-up	~0.8	~0.8	~0.5
1-down	~0.2	~0.2	~0.5
2-down	~0.05	~0.05	~0.05
1-up	~0.7	~0.5	~0.95
1-down	~0.05	~0.05	~0.05
2-up	~0.4	~0.5	~0.05

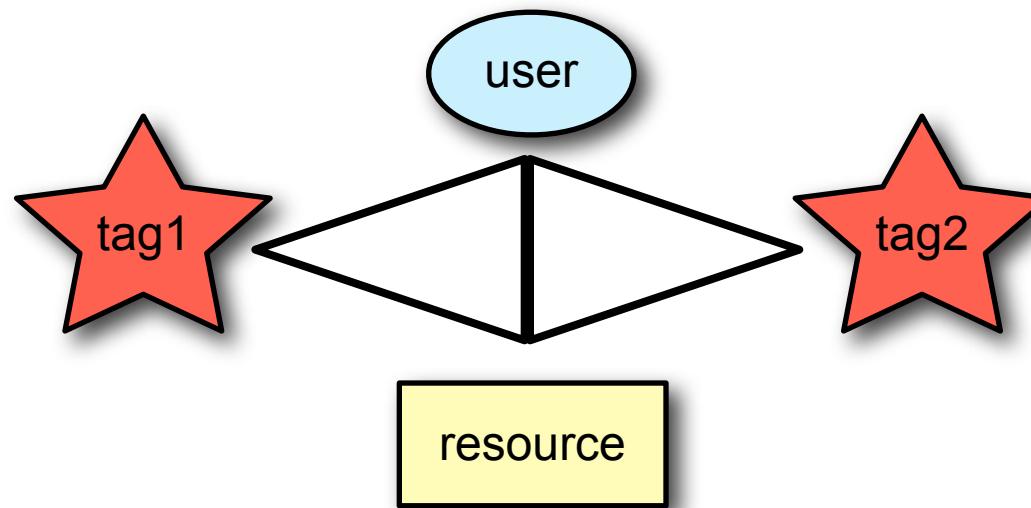
C.Cattuto, D.Benz, A.Hotho, G.Stumme

tag co-occurrence

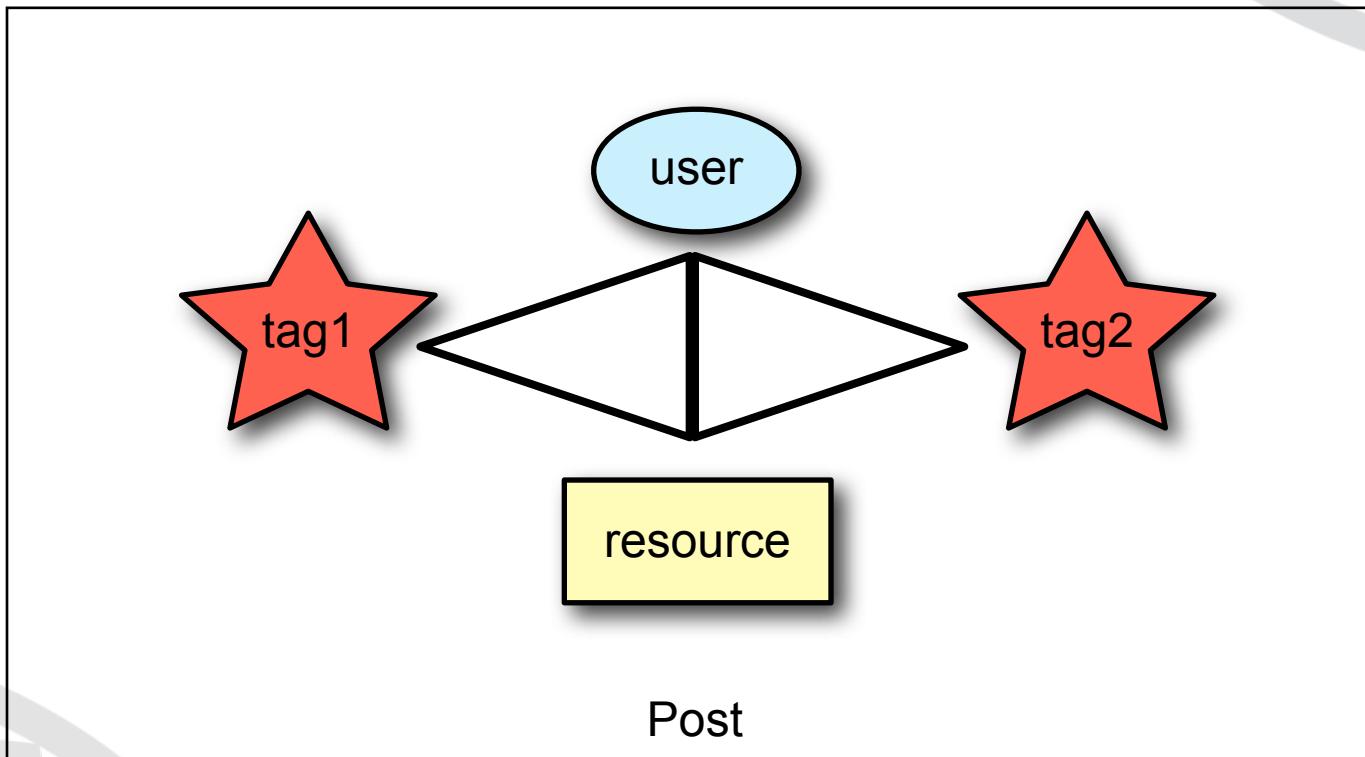
tag co-occurrence



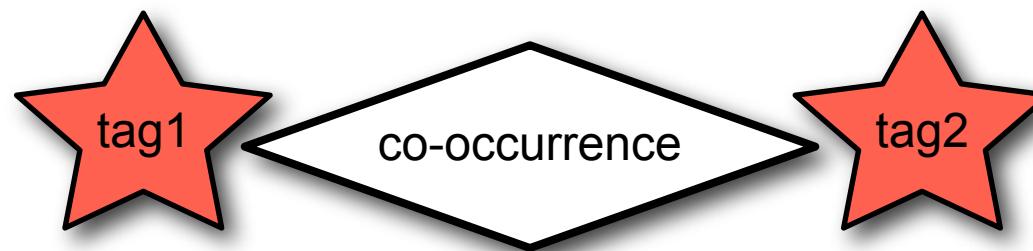
tag co-occurrence



tag co-occurrence

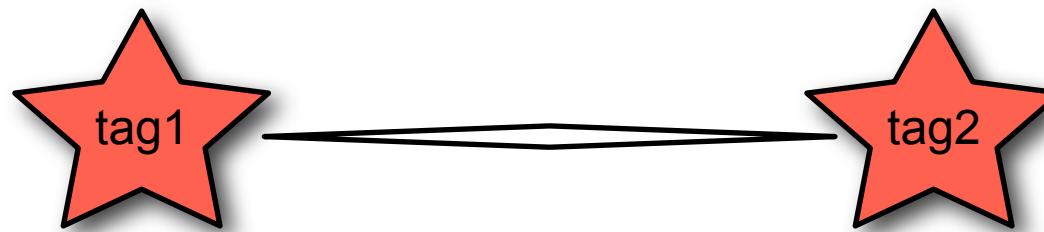


tag co-occurrence

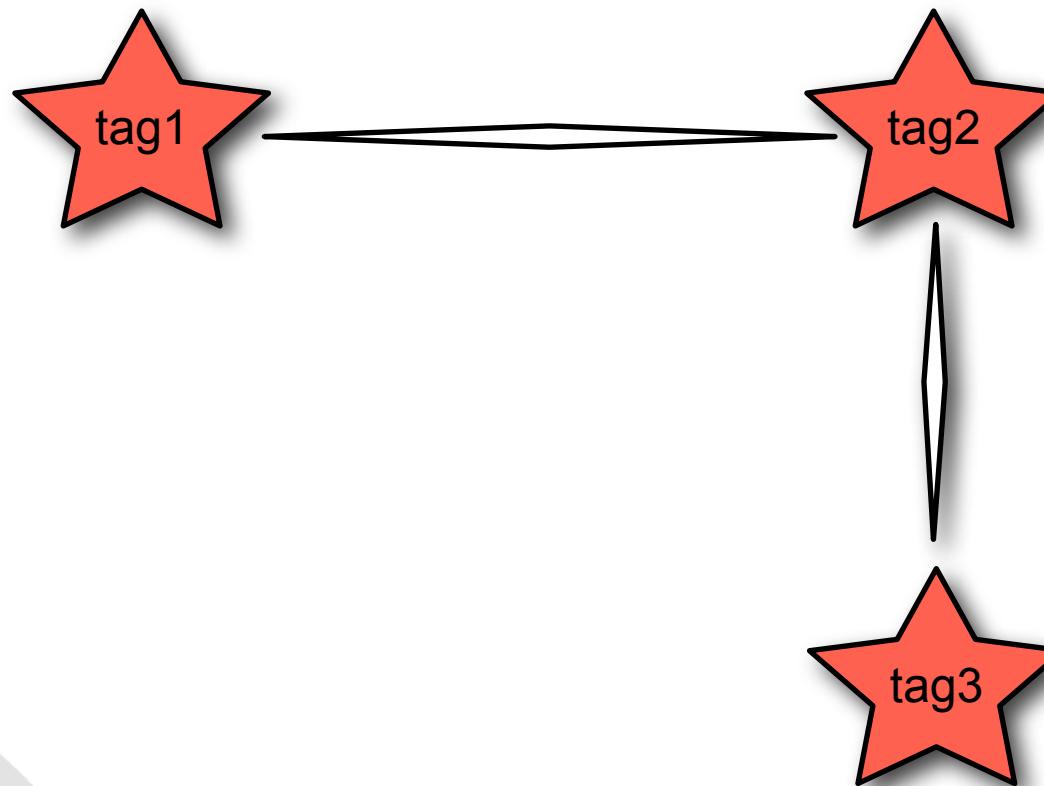


tag co-occurrence network

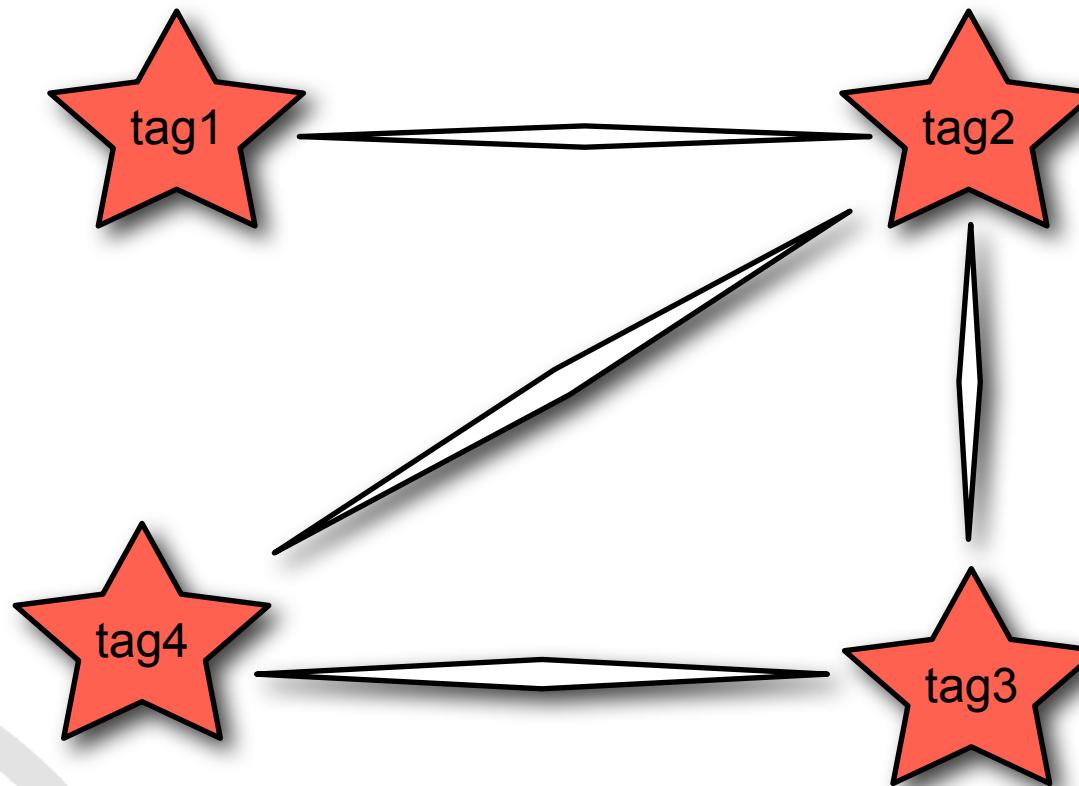
tag co-occurrence network



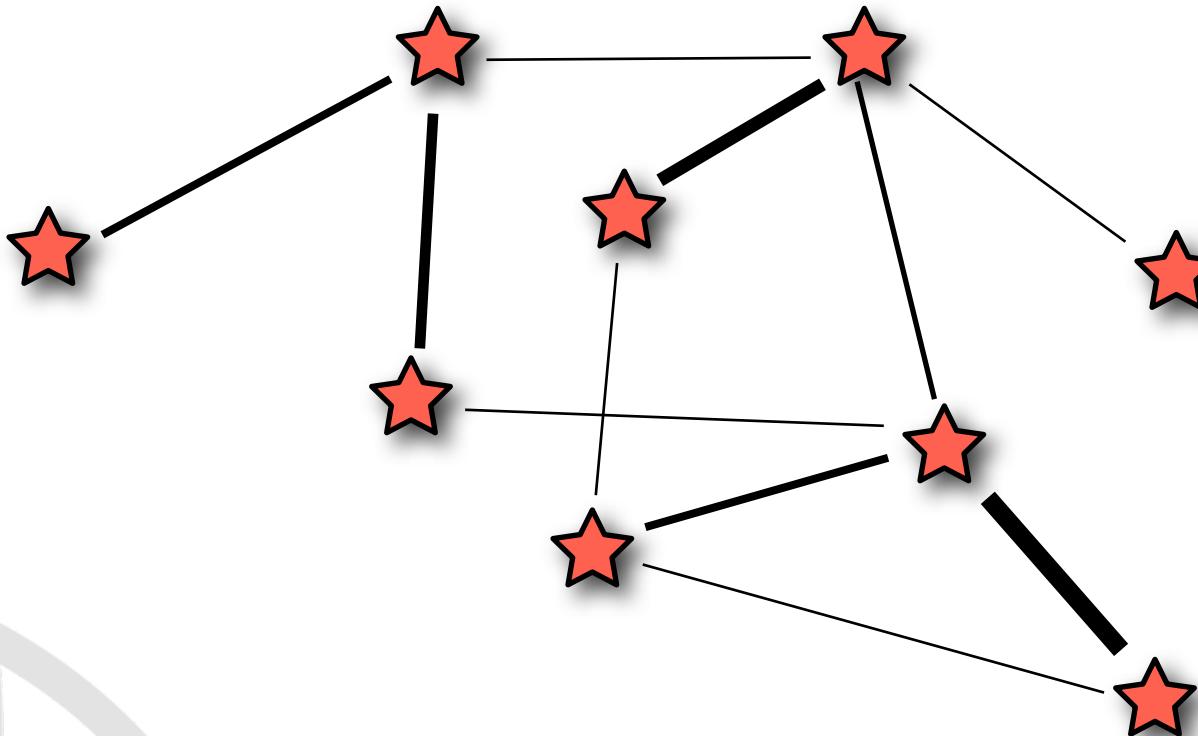
tag co-occurrence network



tag co-occurrence network

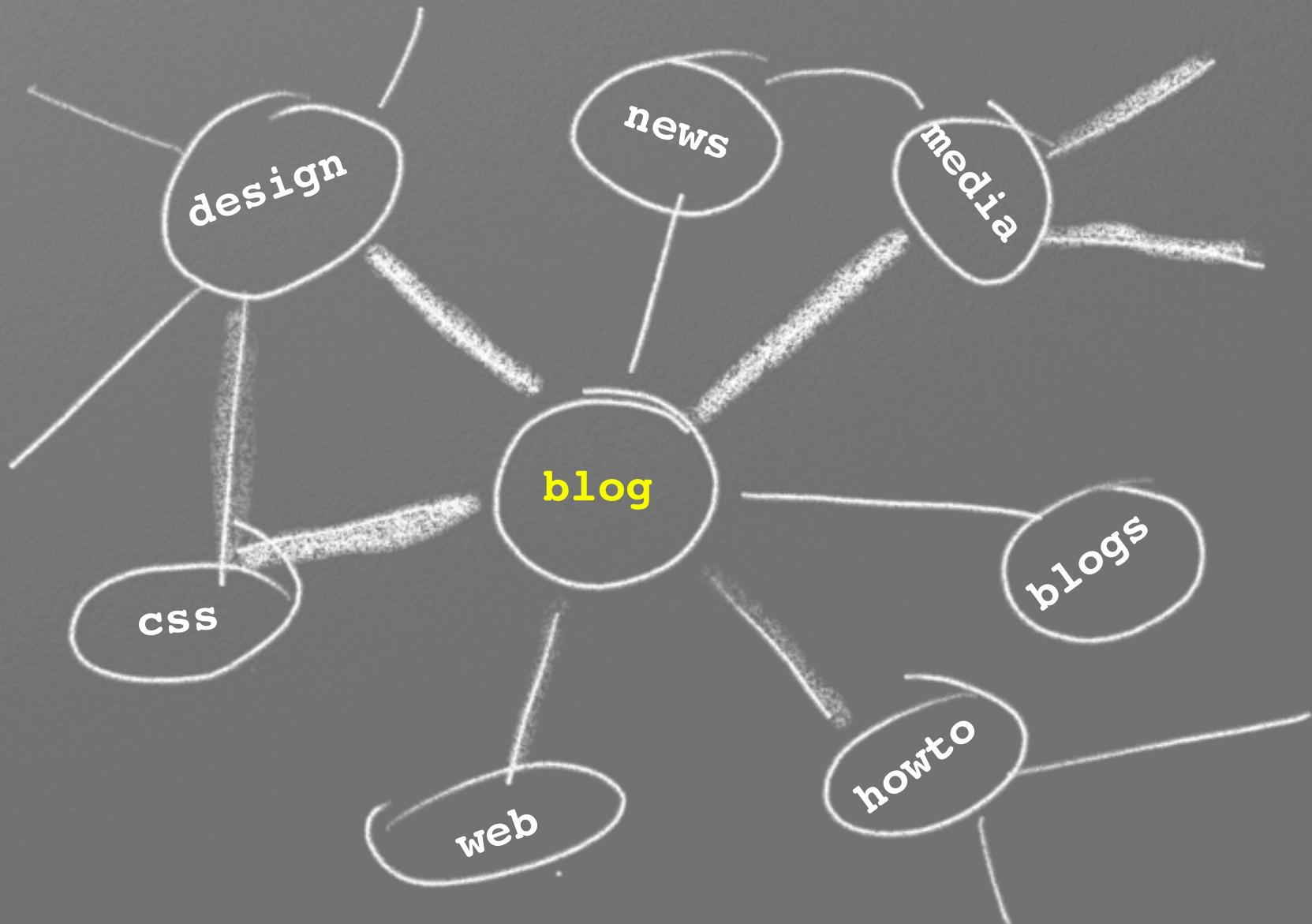


tag co-occurrence network

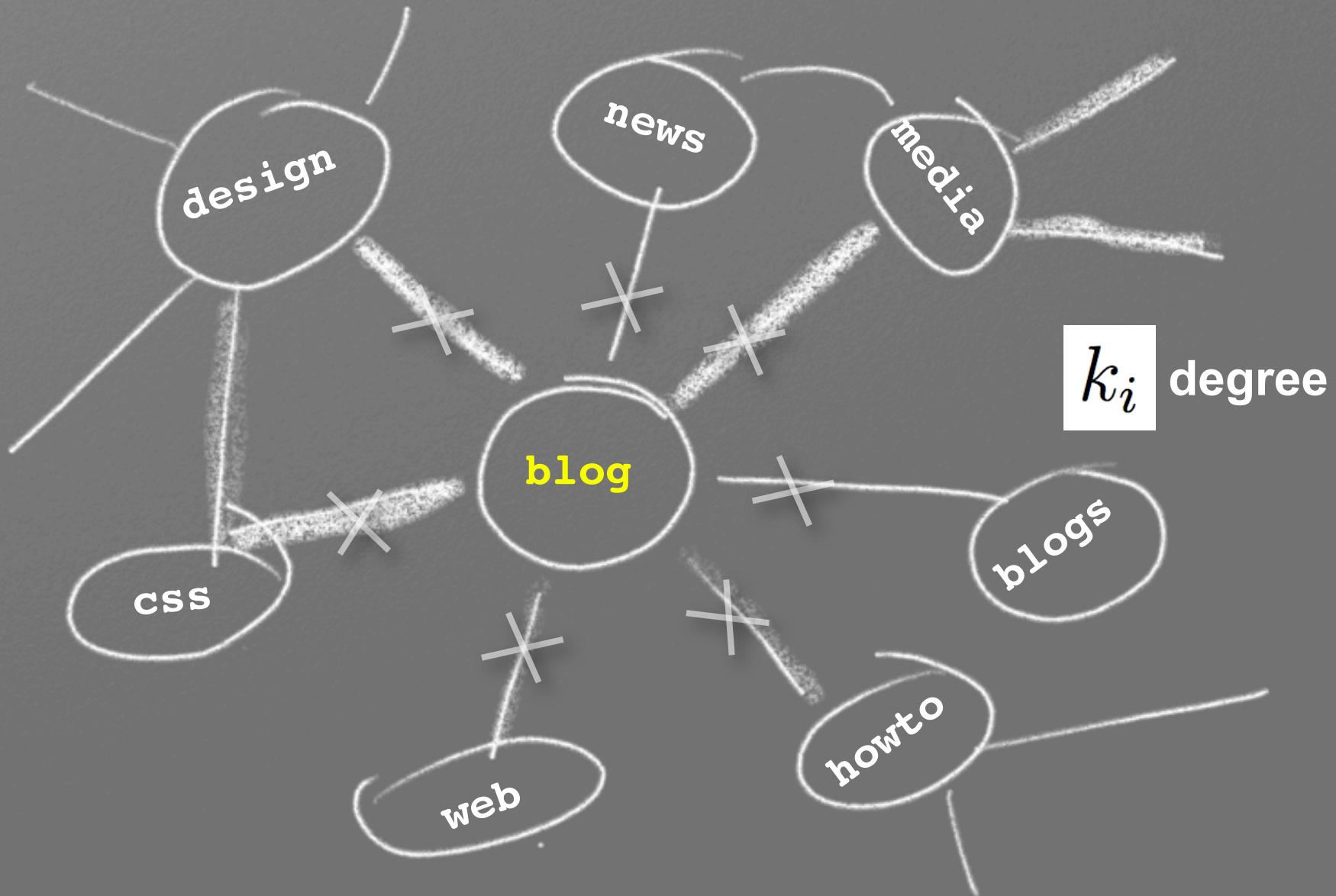


weighted network:
weight = # posts two tags co-occur

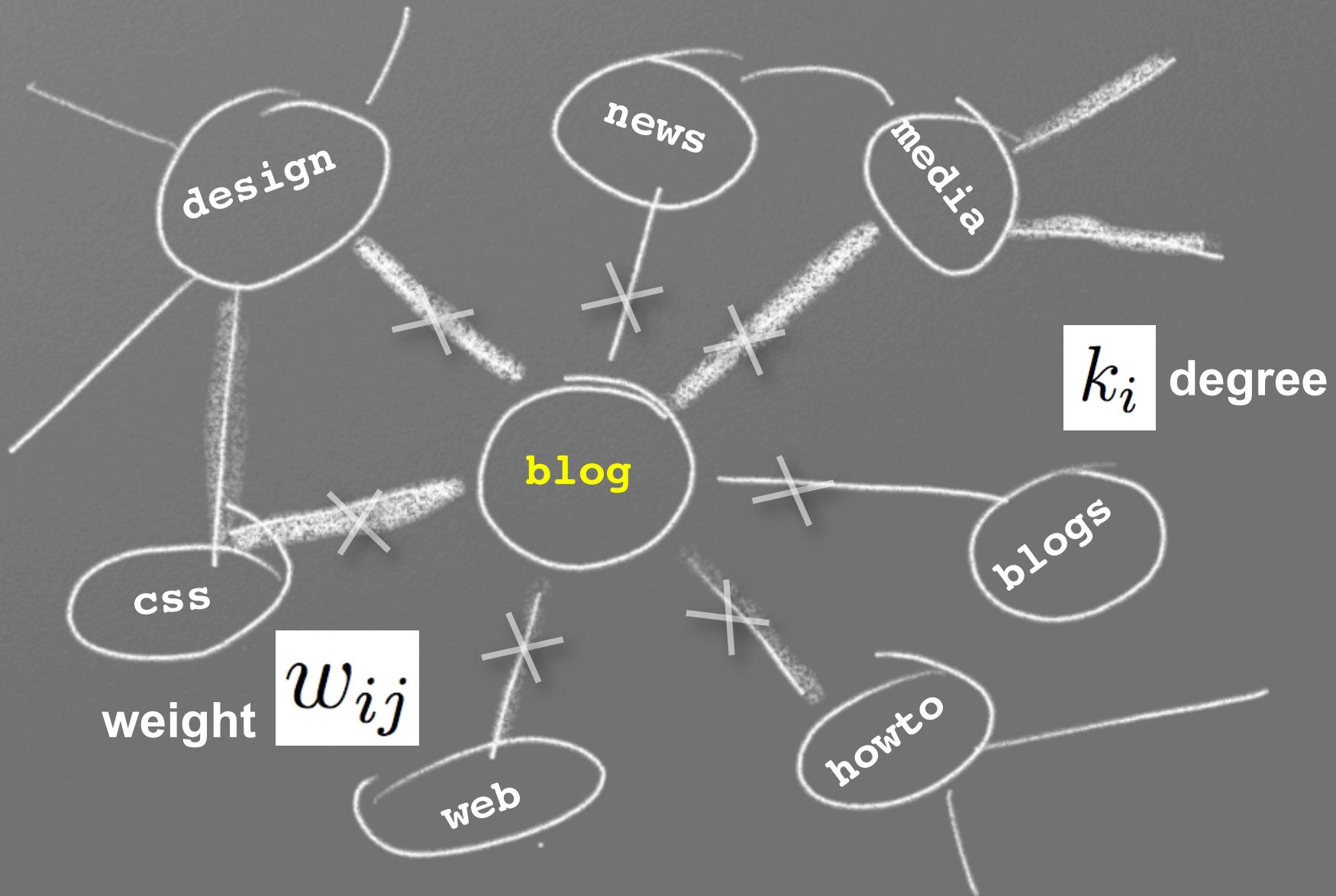
Tagora nodes and links



Tagora nodes and links



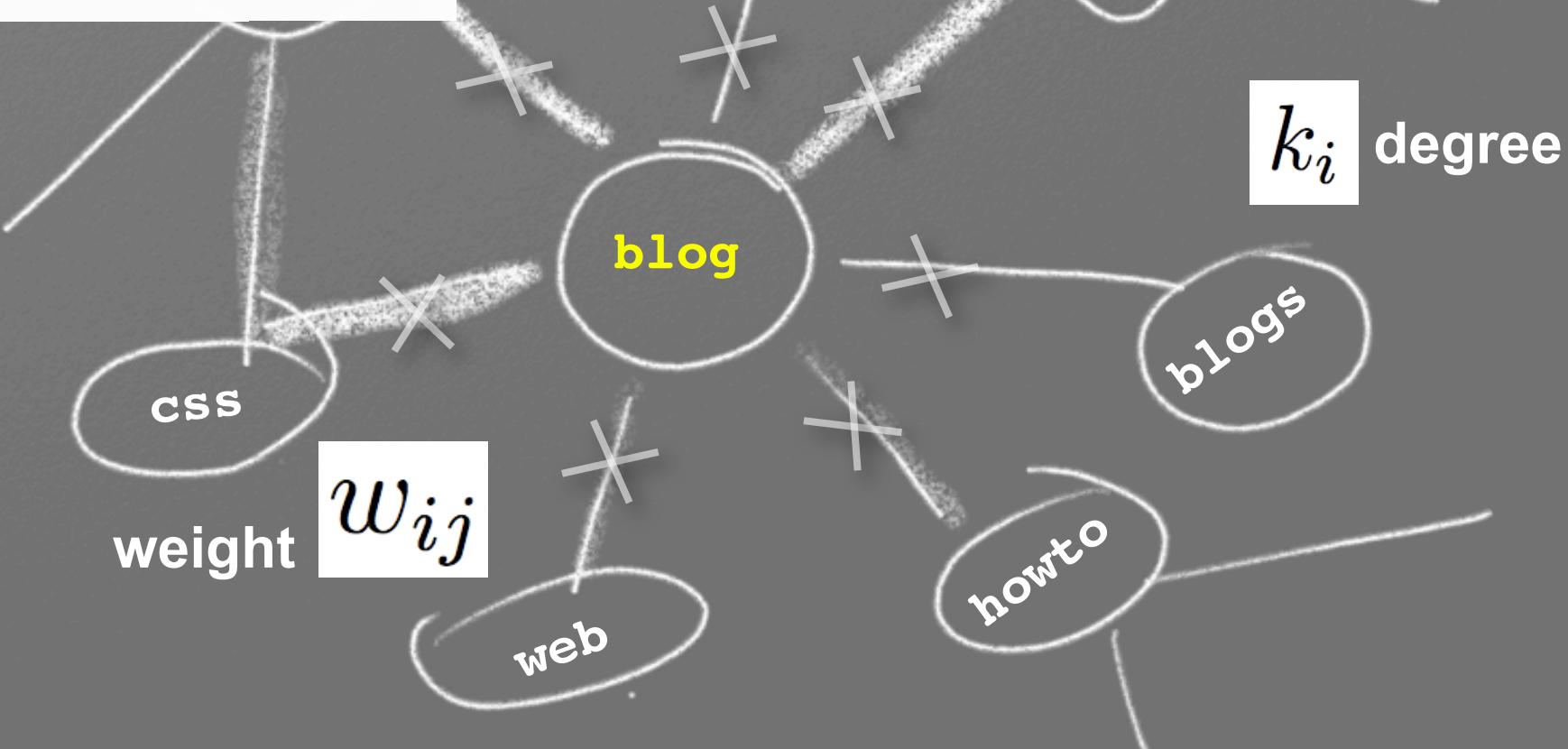
Tagora nodes and links



Tagora nodes and links

strength

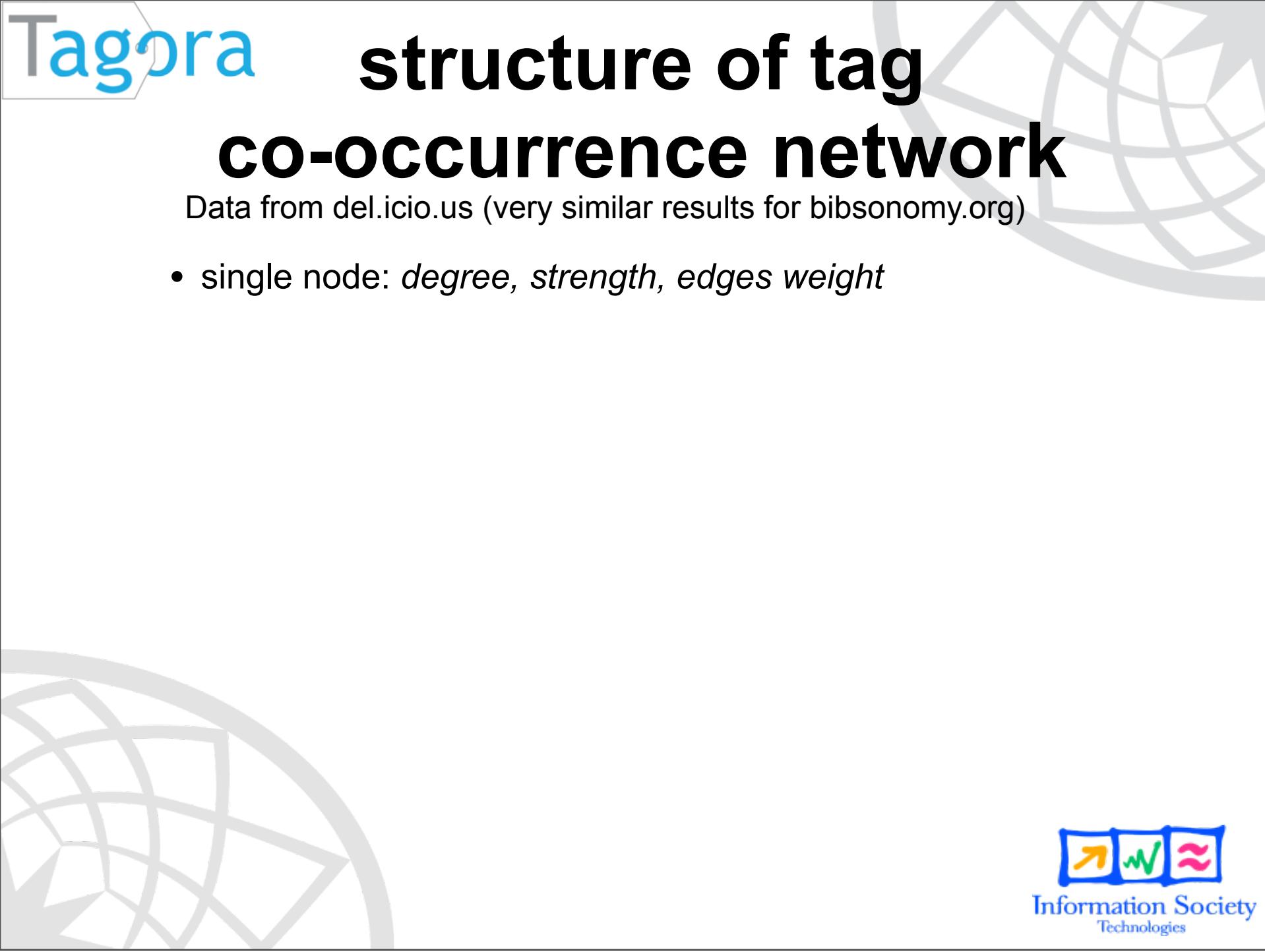
$$s_i = \sum_{j \in \mathcal{V}(i)} w_{ij}$$



k_i degree

weight

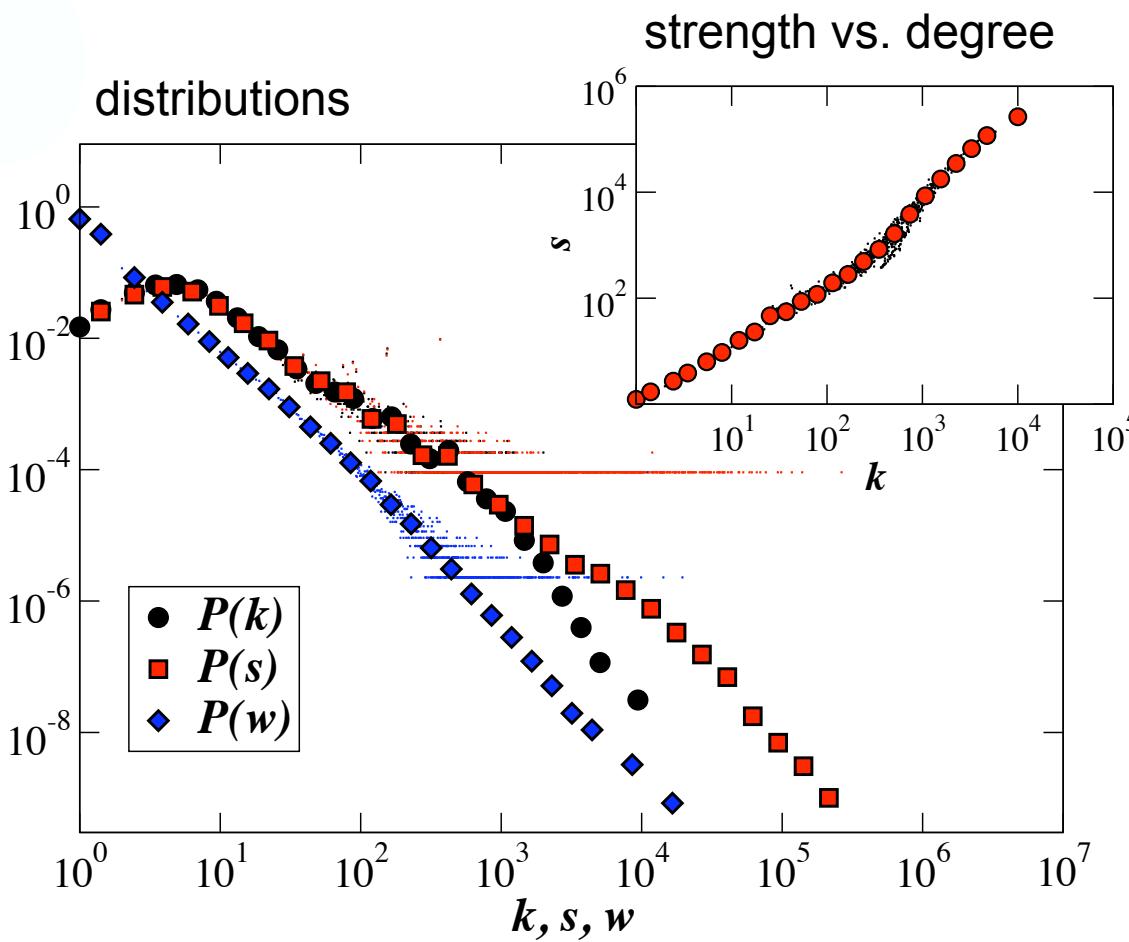
$$w_{ij}$$



structure of tag co-occurrence network

Data from del.icio.us (very similar results for bibsonomy.org)

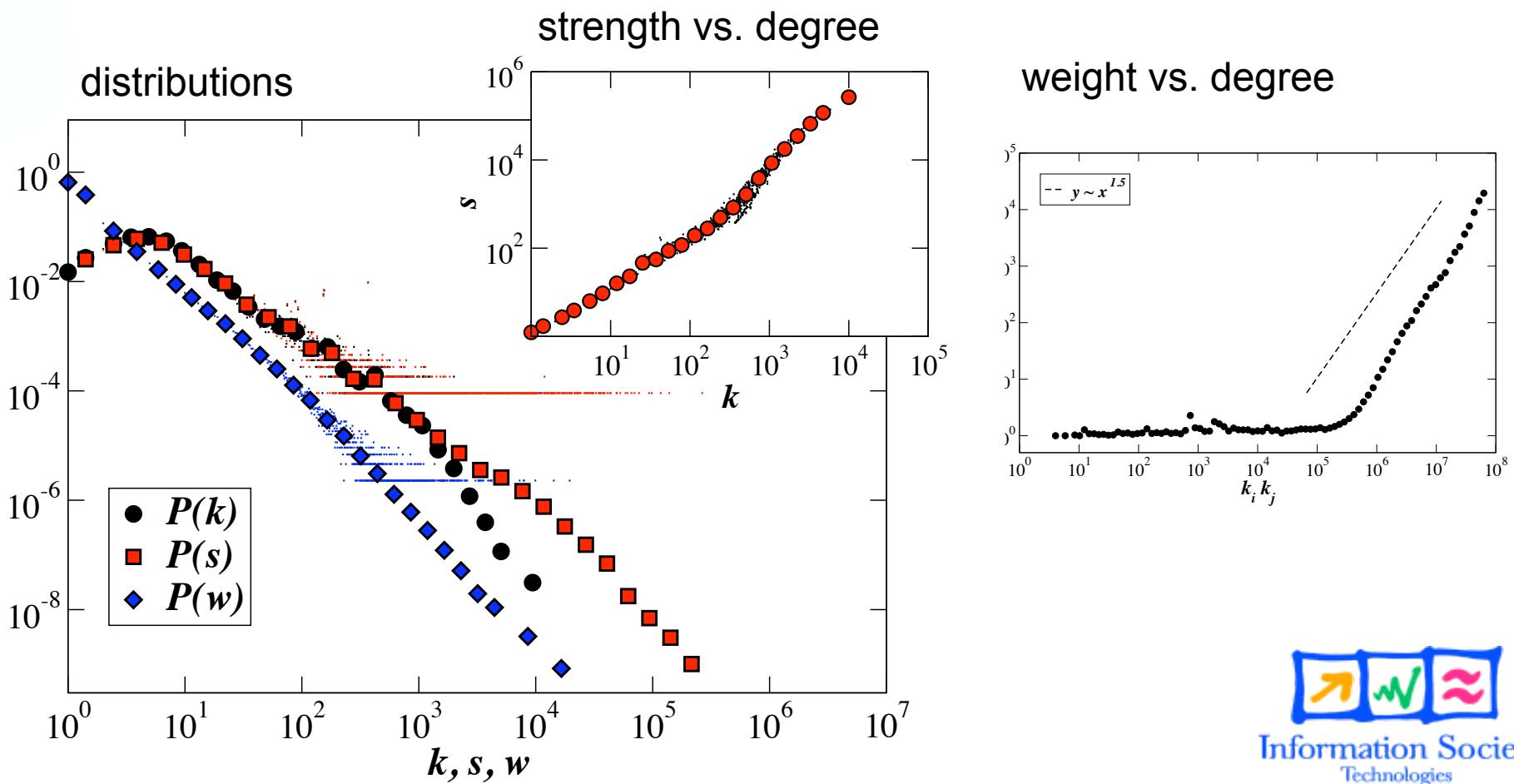
- single node: *degree, strength, edges weight*



structure of tag co-occurrence network

Data from del.icio.us (very similar results for bibsonomy.org)

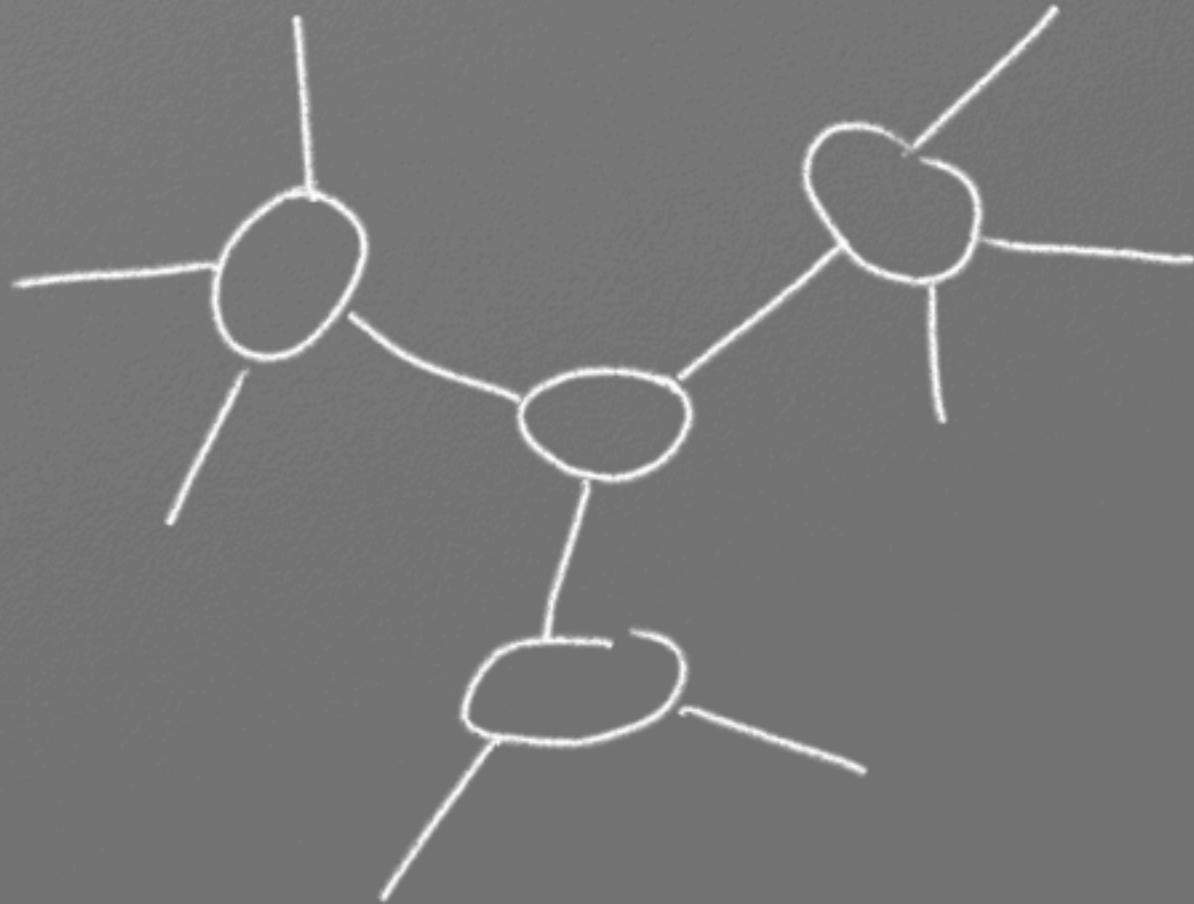
- single node: *degree, strength, edges weight*



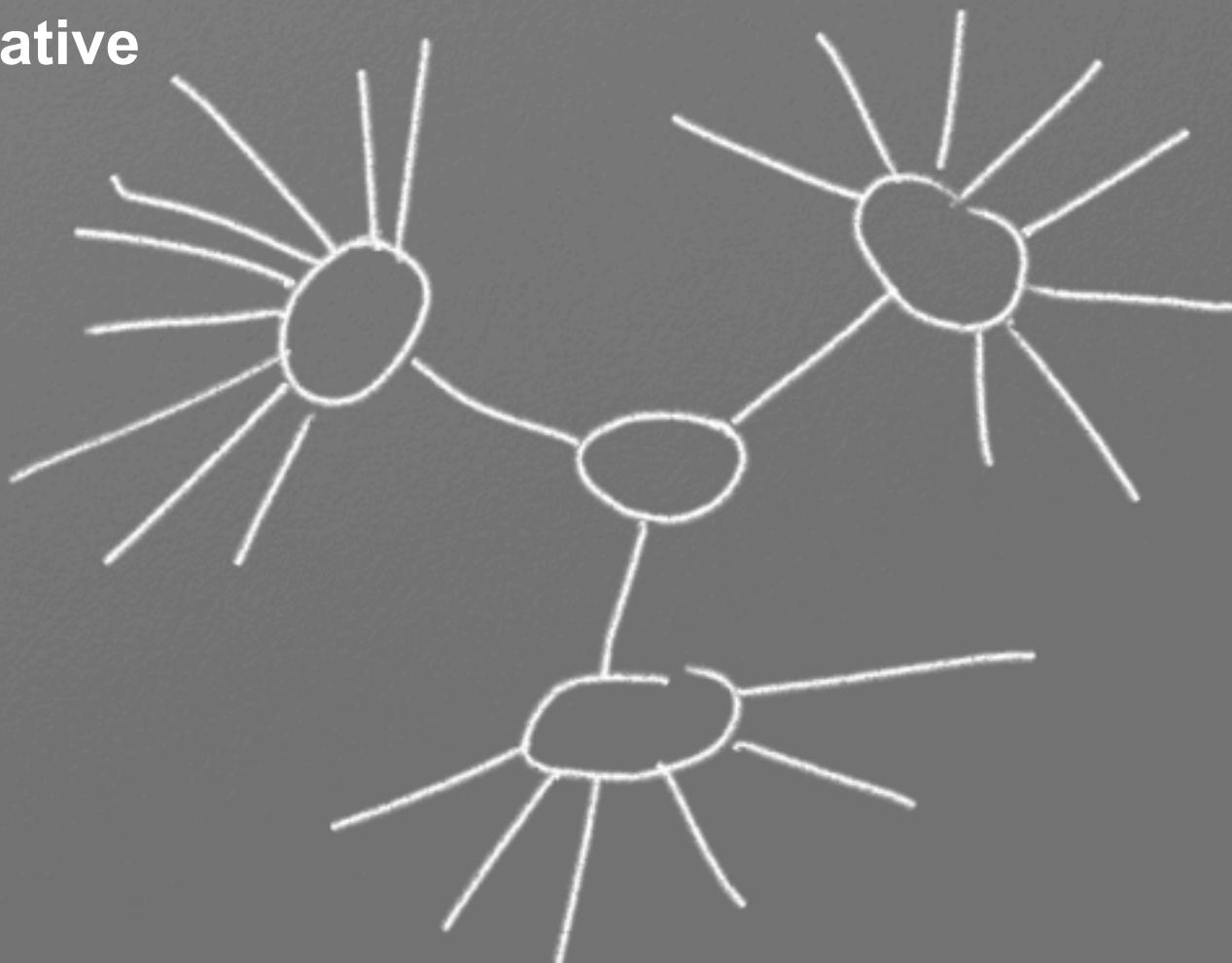
Tagora network topology

Tagora network topology

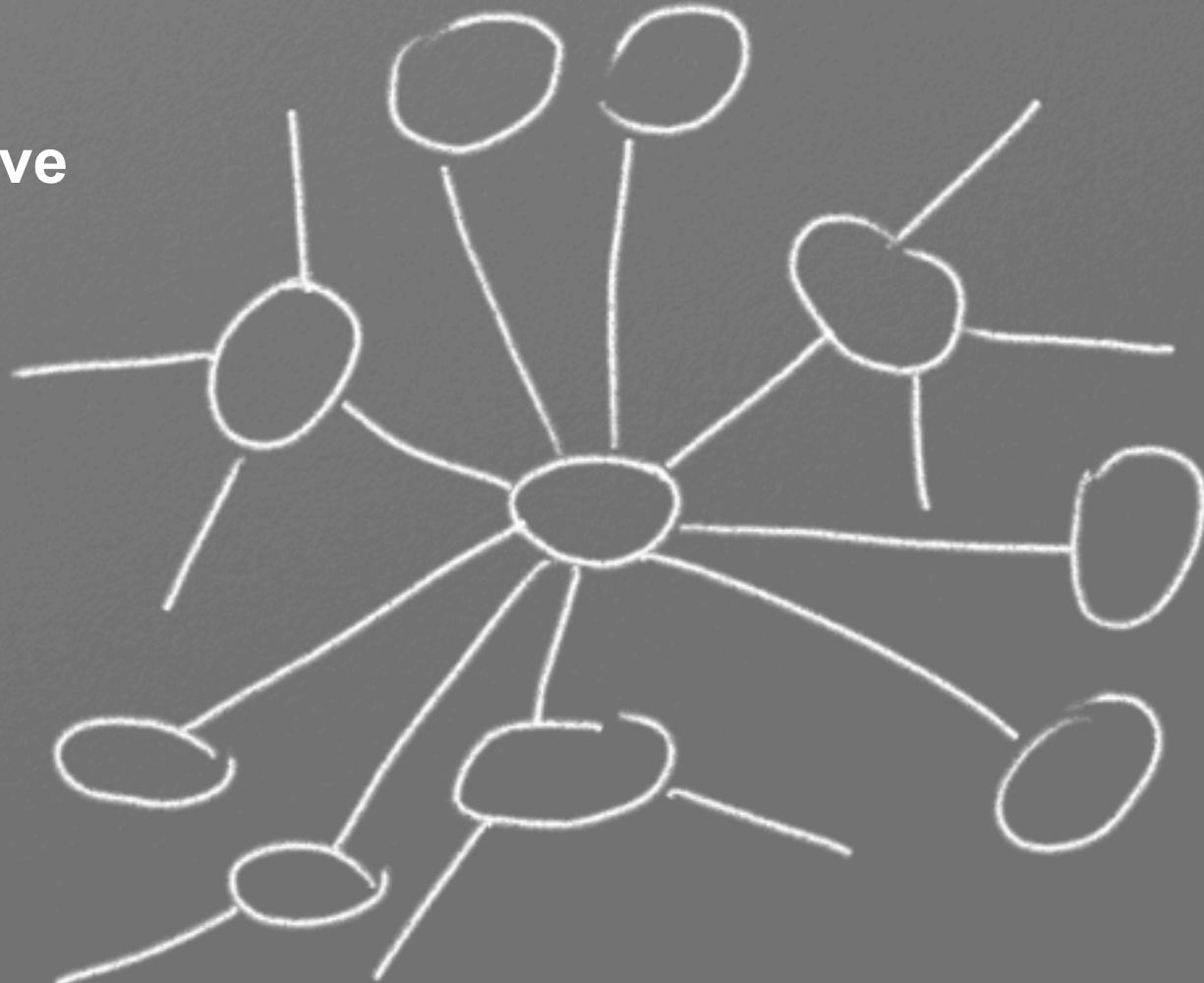
assortative



disassortative



disassortative



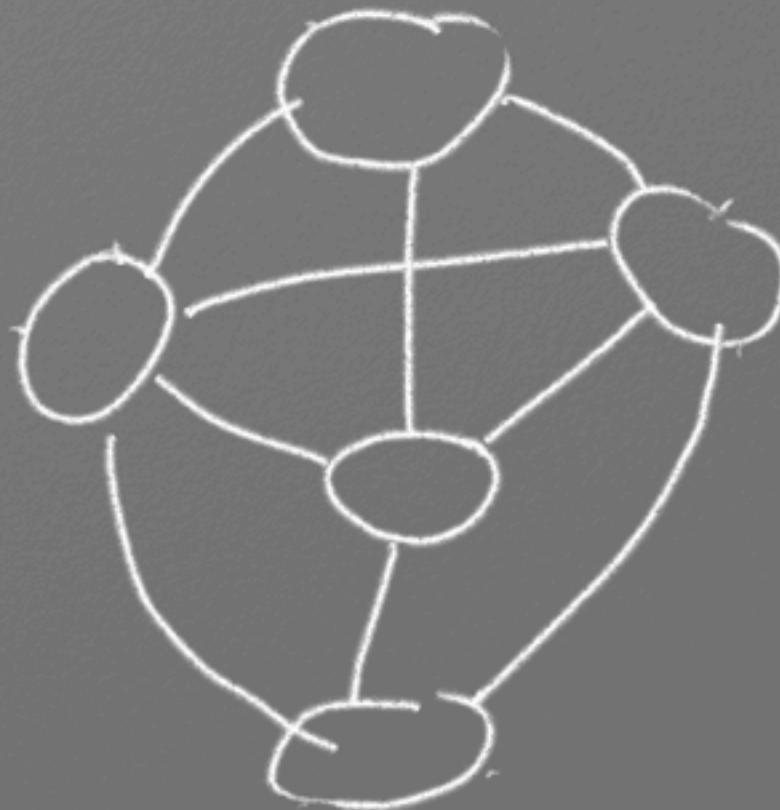
Tagora network topology

low clustering

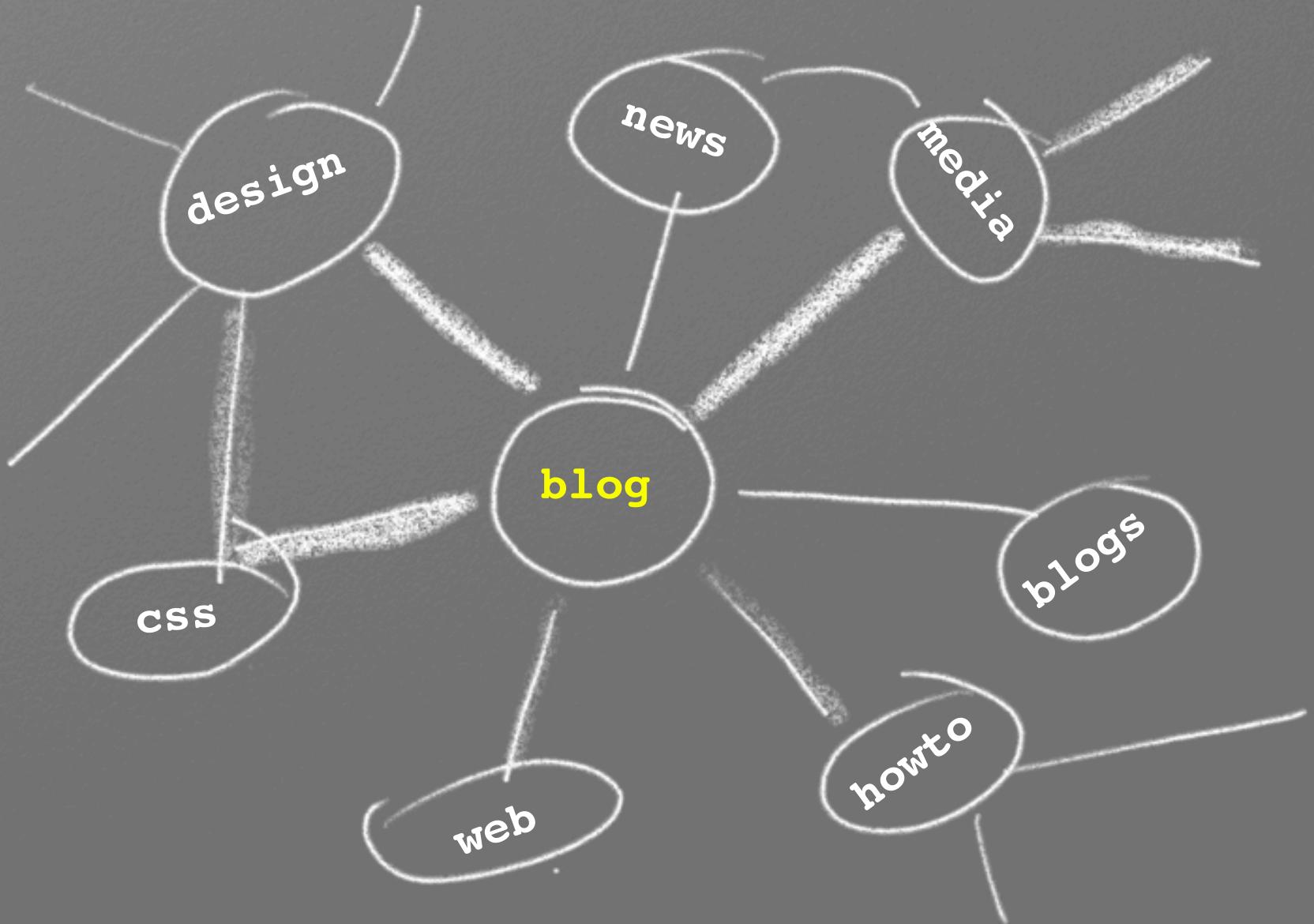


Tagora network topology

high clustering

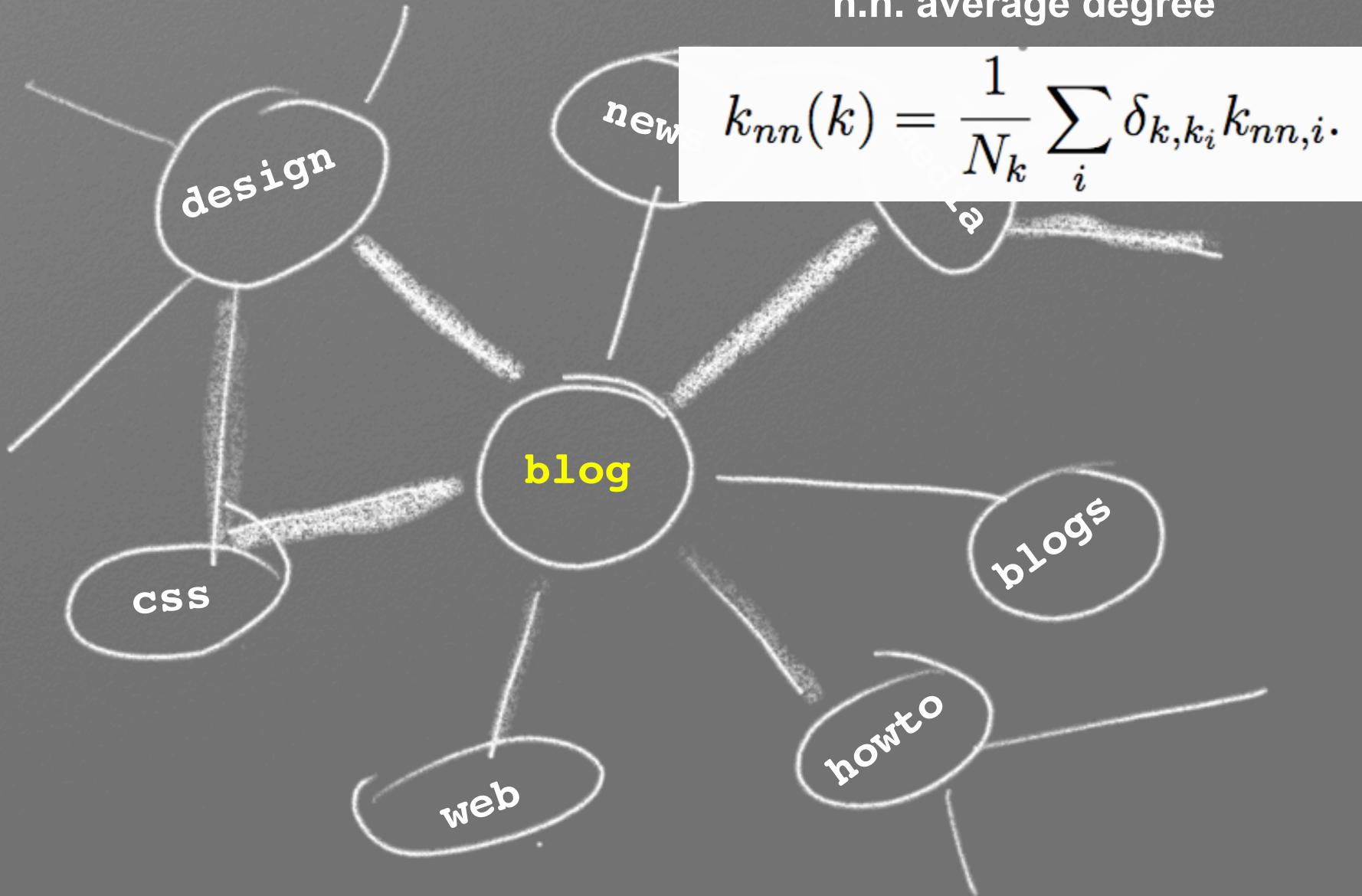


Tagora clustering and assortativity



Tagora clustering and assortativity

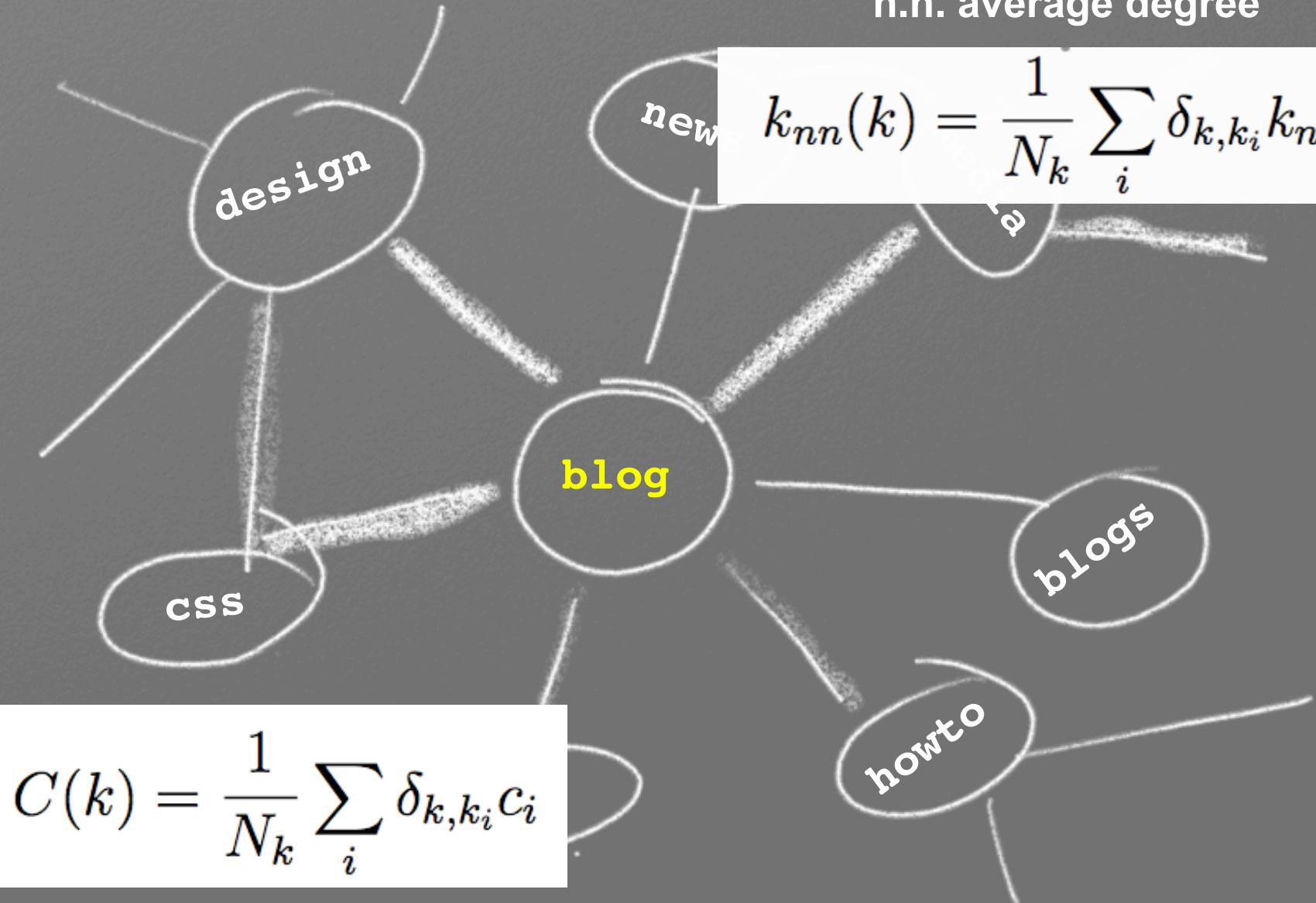
n.n. average degree



Tagora clustering and assortativity

n.n. average degree

$$k_{nn}(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} k_{nn,i}.$$



$$C(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} c_i$$

clustering coefficient

Tagora clustering and assortativity

n.n. average degree

$$k_{nn}(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} k_{nn,i}.$$

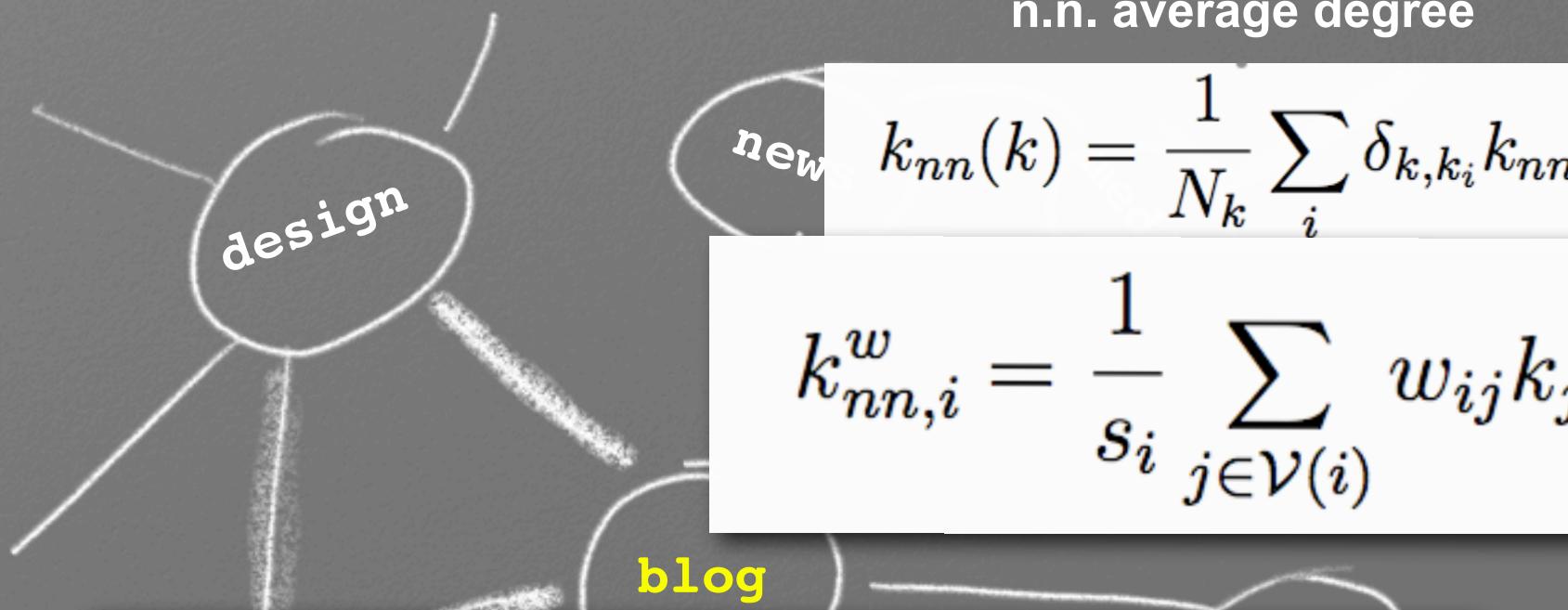
$$k_{nn,i}^w = \frac{1}{s_i} \sum_{j \in \mathcal{V}(i)} w_{ij} k_j$$

$$C(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} c_i$$

clustering coefficient

Tagora clustering and assortativity

n.n. average degree


$$k_{nn}(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} k_{nn,i}.$$

$$k_{nn,i}^w = \frac{1}{s_i} \sum_{j \in \mathcal{V}(i)} w_{ij} k_j$$

$$c^w(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,h \in \mathcal{V}(i), j \in \mathcal{V}(h)} \frac{(w_{ij} + w_{ih})}{2}$$

$$C(k) = \frac{1}{N_k} \sum_i \delta_{k,k_i} c_i$$

clustering coefficient

Tagora correlations in tag co-occurrence network

Data from del.icio.us (very similar results for bibsonomy.org)

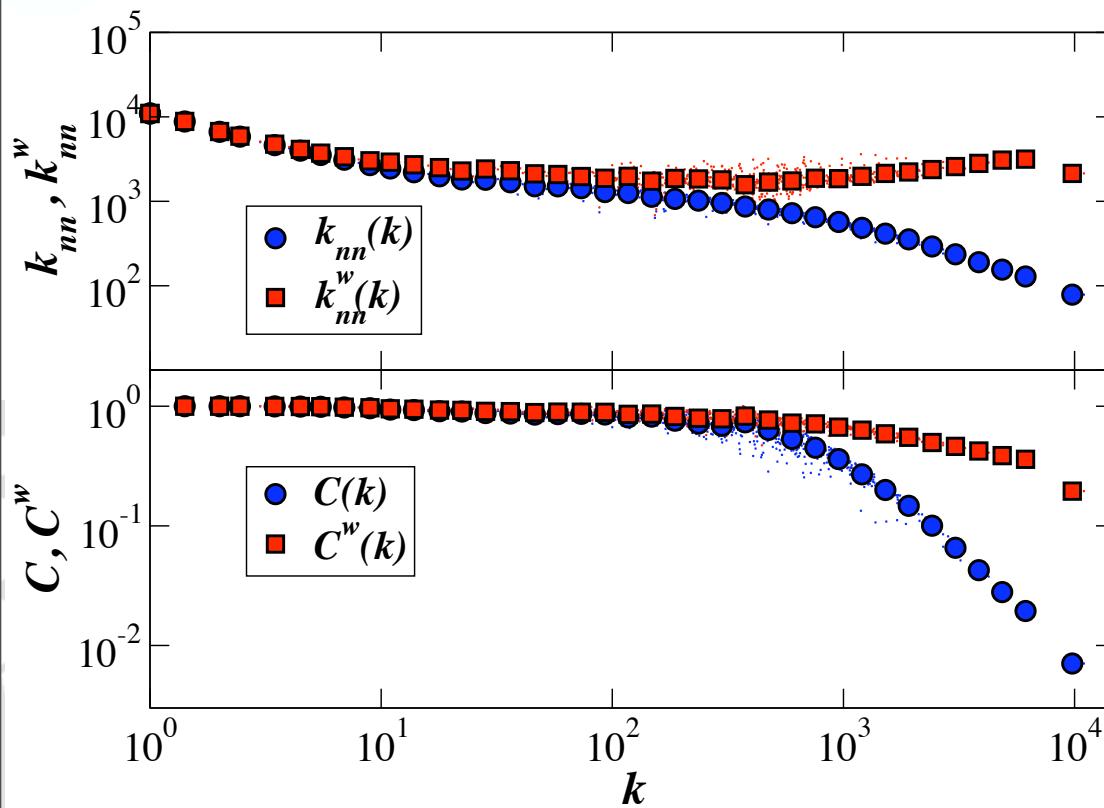


Information Society
Technologies

Tagora correlations in tag co-occurrence network

Data from del.icio.us (very similar results for bibsonomy.org)

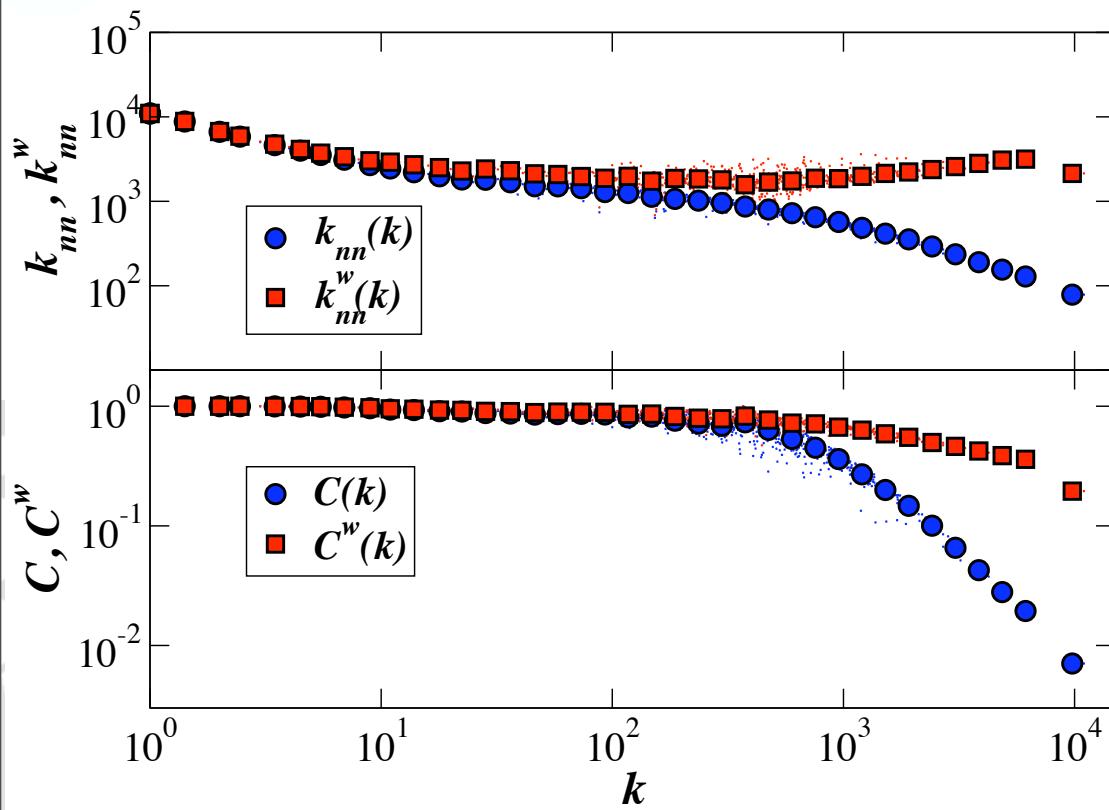
- Average nearest neighbor degree
- Clustering coefficients



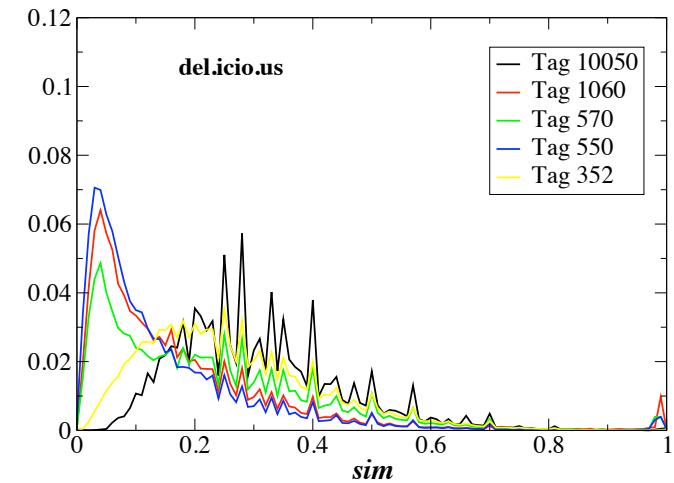
Tagora correlations in tag co-occurrence network

Data from del.icio.us (very similar results for bibsonomy.org)

- Average nearest neighbor degree
- Clustering coefficients



Distribution of pair-wise structural similarities



Information Society
Technologies

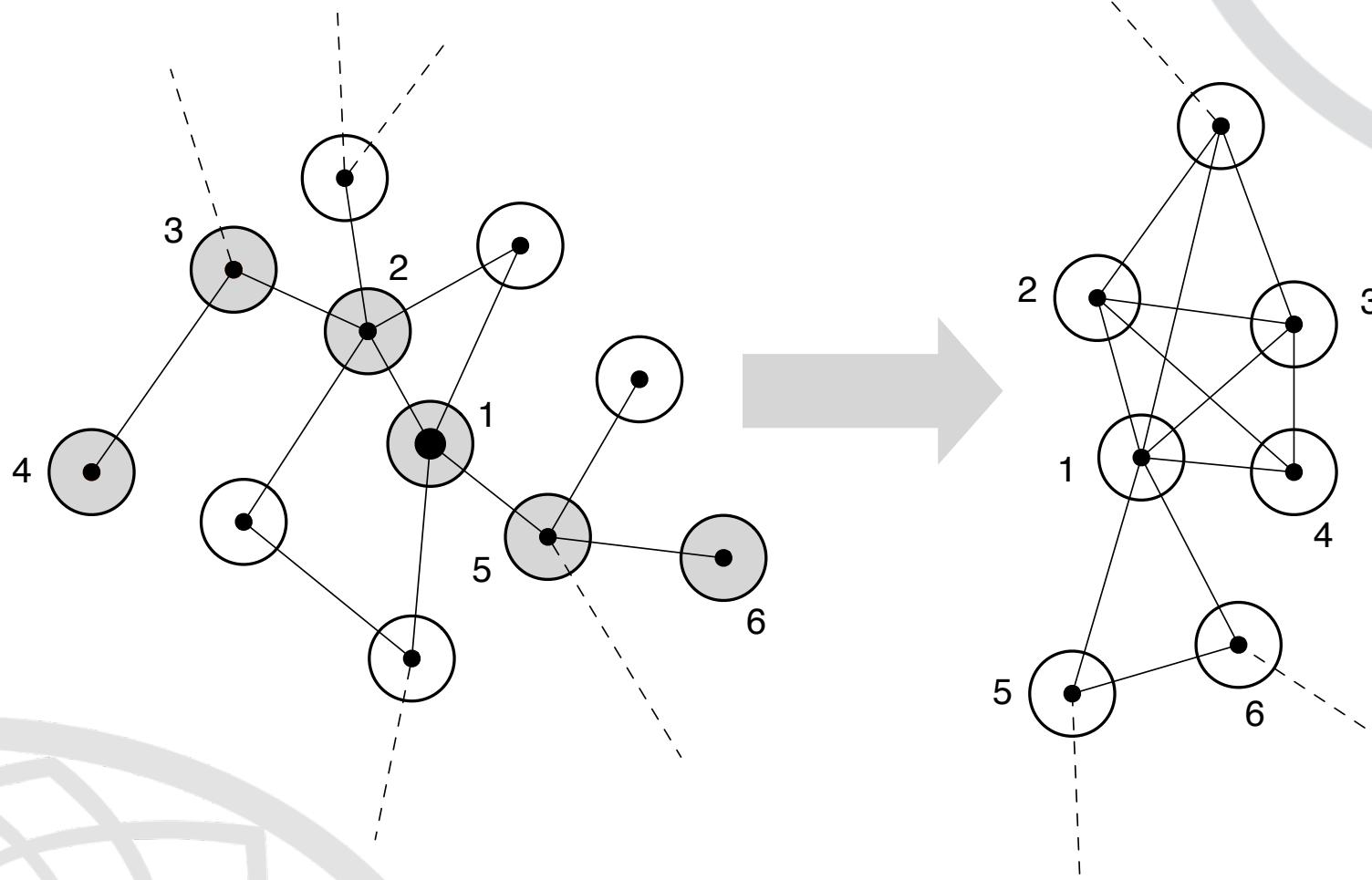
Modeling



Modeling



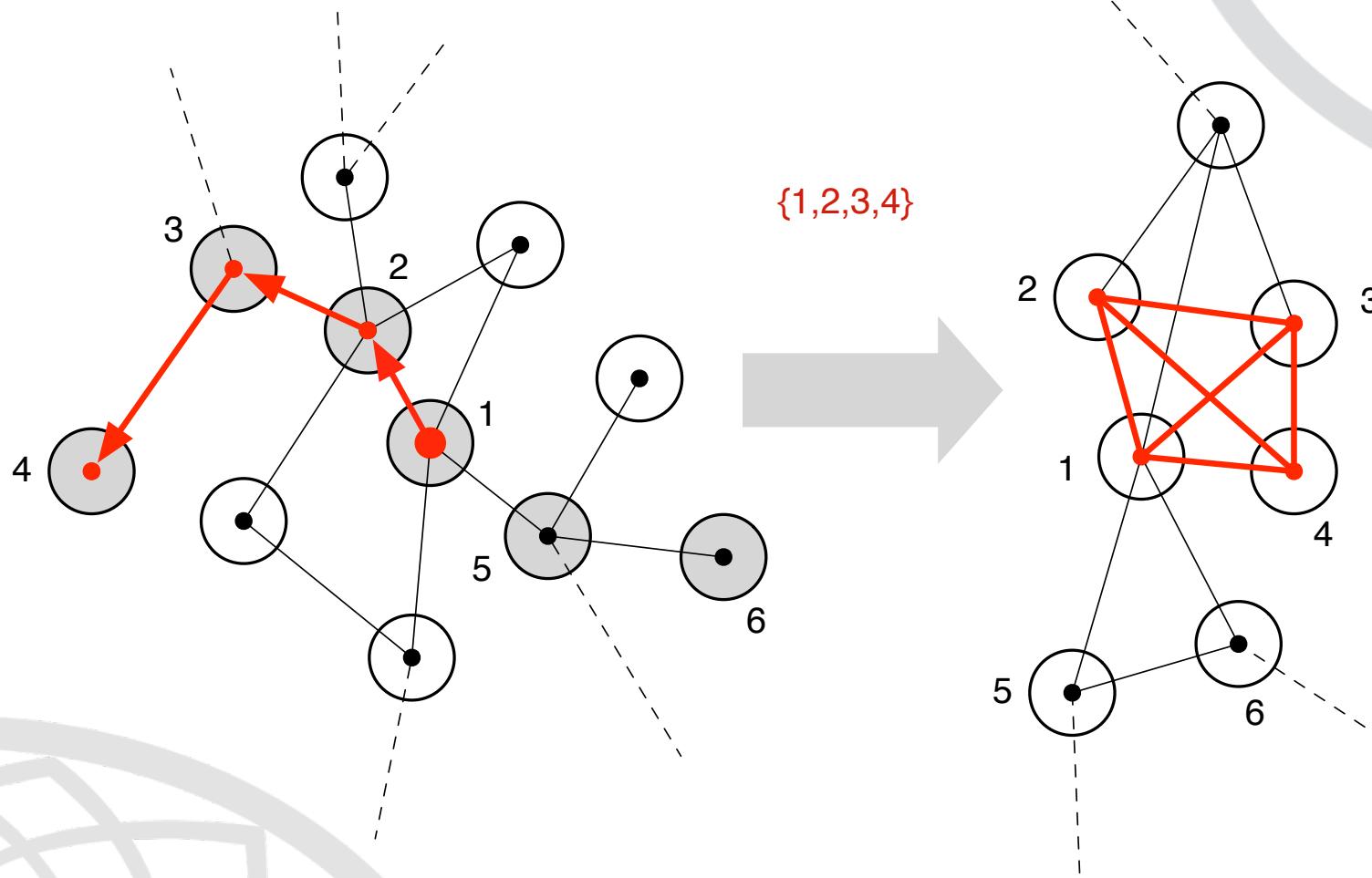
model idea



shared semantic graph (latent)

tag co-occurrence network

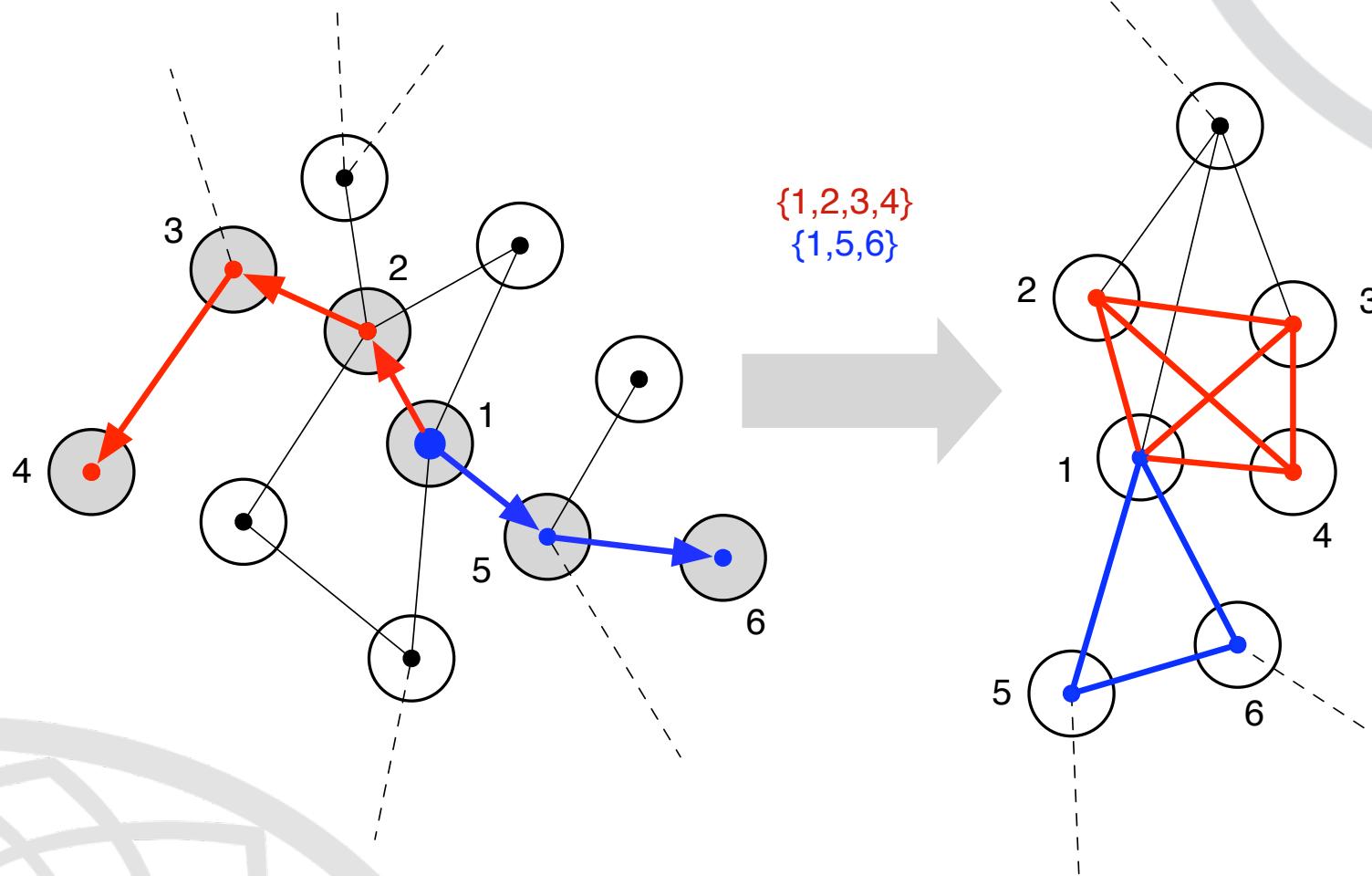
model idea



shared semantic graph (latent)

tag co-occurrence network

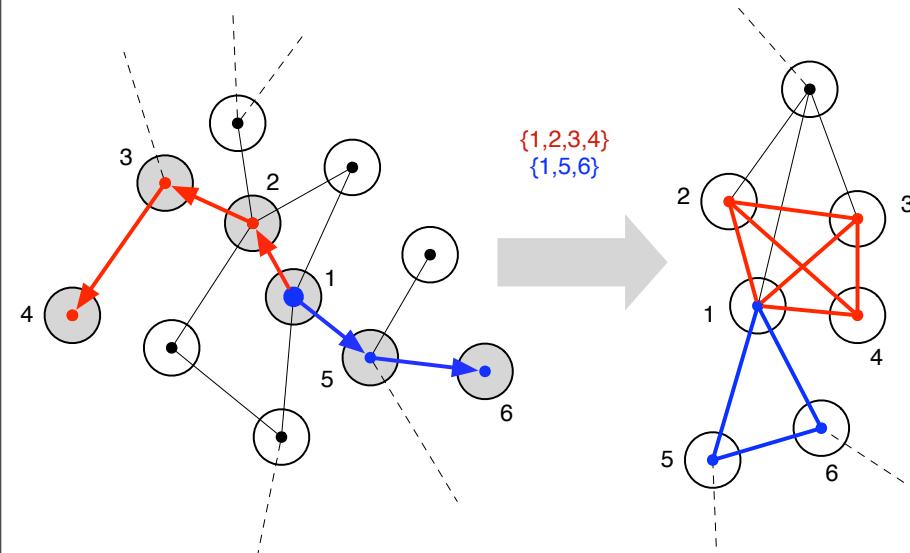
model idea



shared semantic graph (latent)

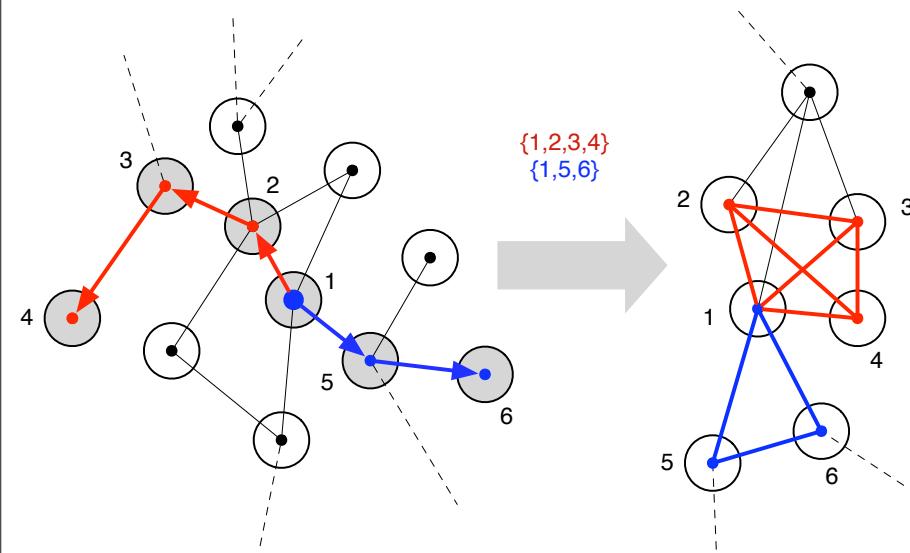
tag co-occurrence network

Model ingredients



Model ingredients

1. semantic graph topology

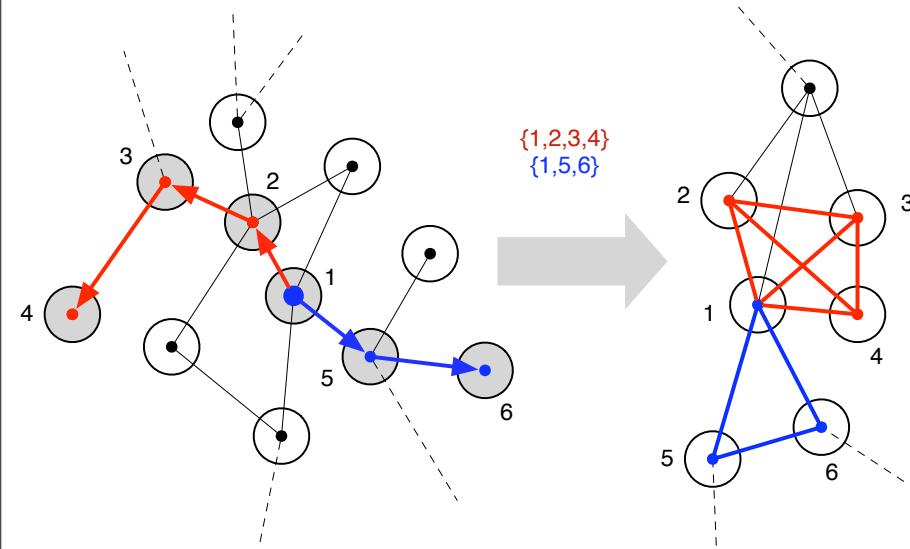


Model ingredients

1. semantic graph topology

networks considered

- Watts and Strogatz (*small world*)
- random scale free (configuration model)
- Erdös-Renyi
- ...



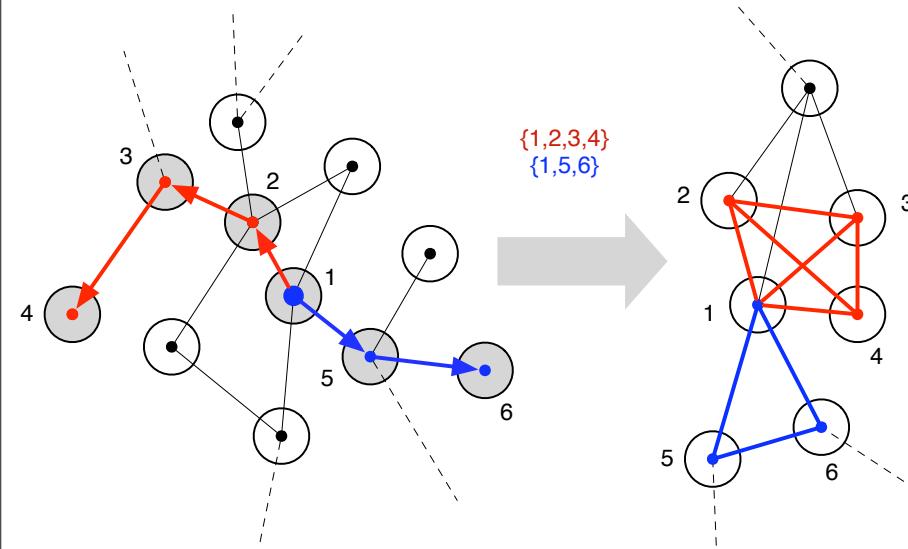
Model ingredients

1. semantic graph topology

2. length of random walks

networks considered

- Watts and Strogatz (*small world*)
- random scale free (configuration model)
- Erdös-Renyi
- ...

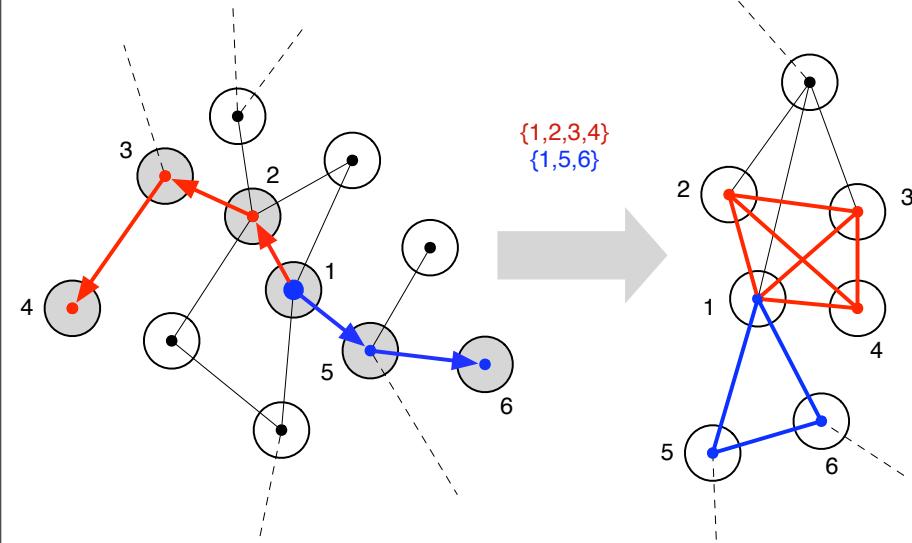


Model ingredients

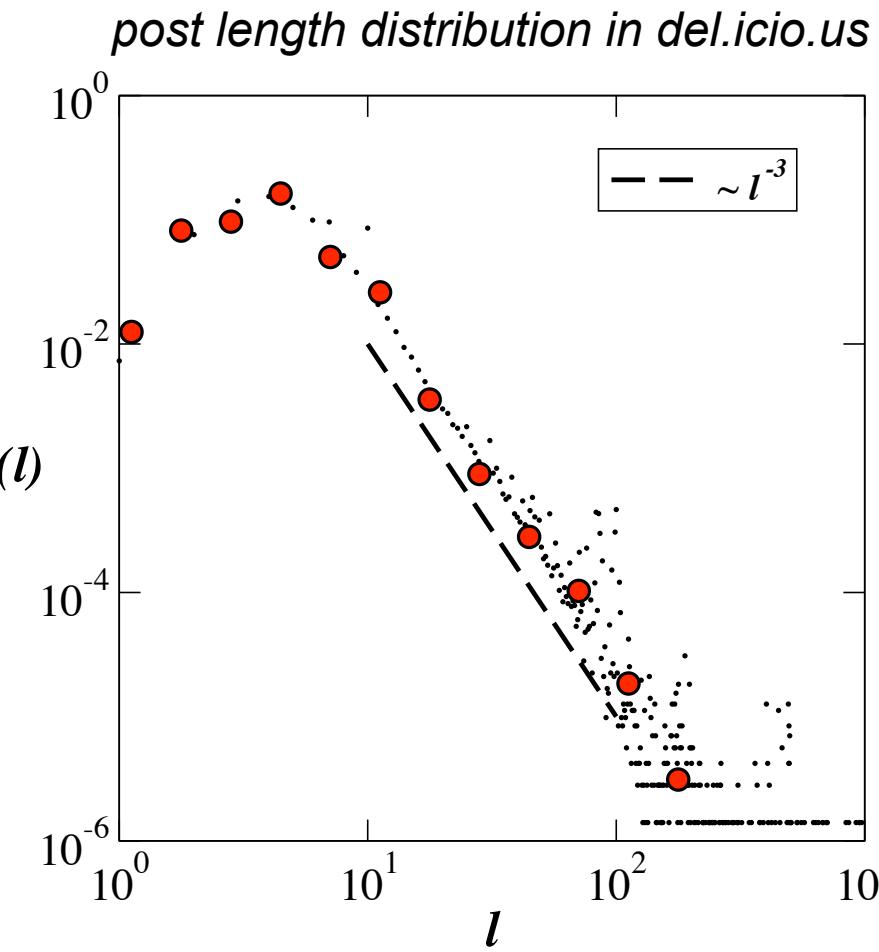
1. semantic graph topology

networks considered

- Watts and Strogatz (*small world*)
- random scale free (configuration model)
- Erdős-Renyi
- ...

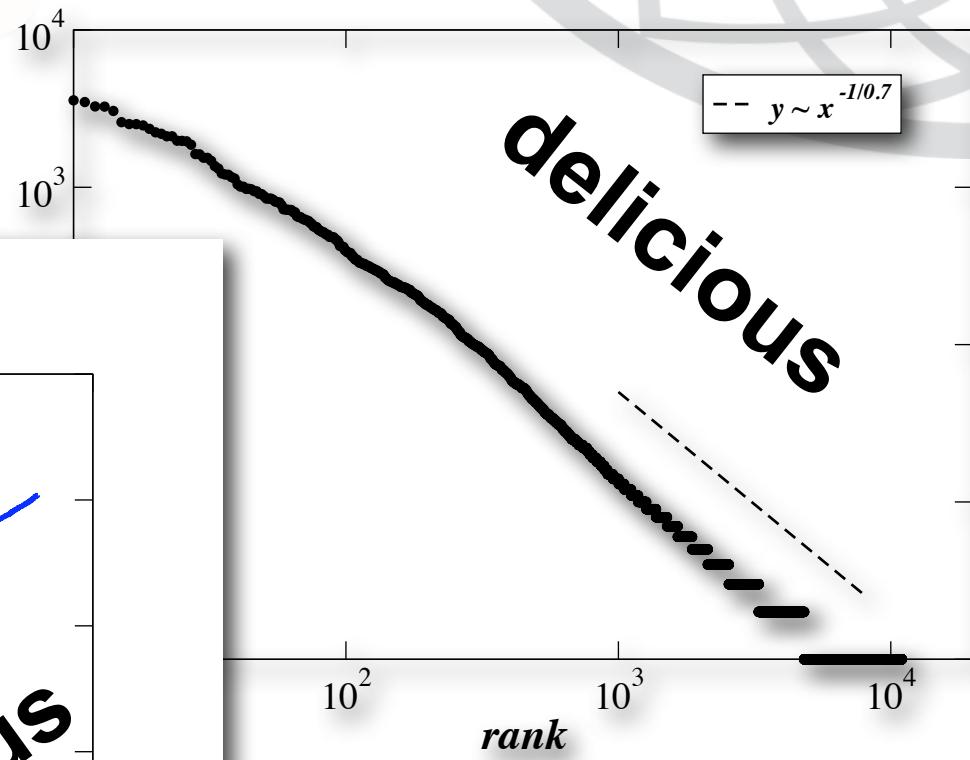
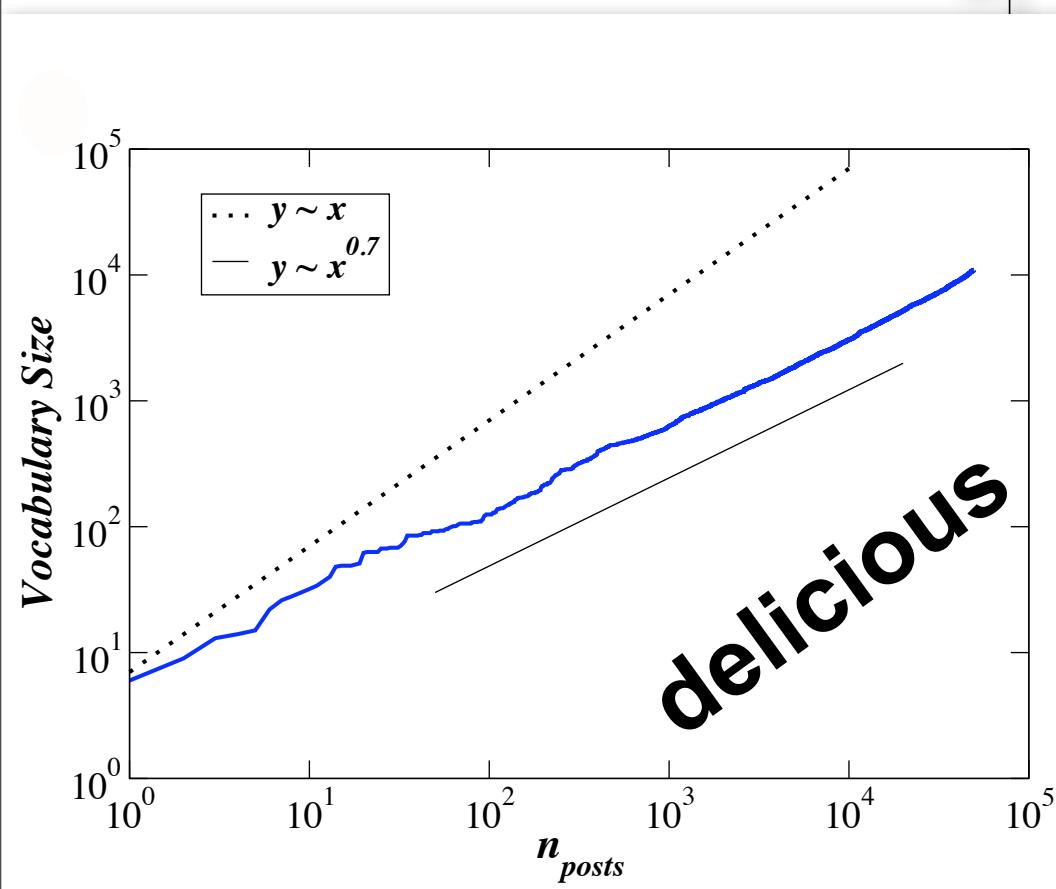


2. length of random walks



Tagora comparison with real systems

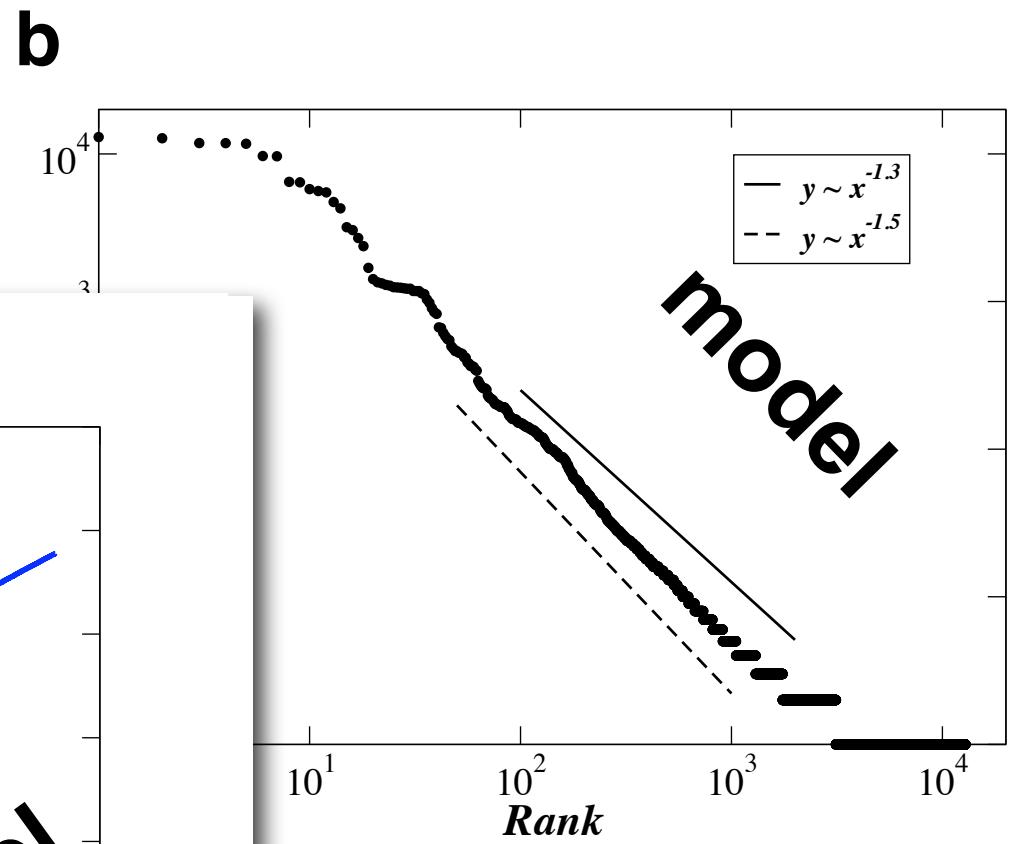
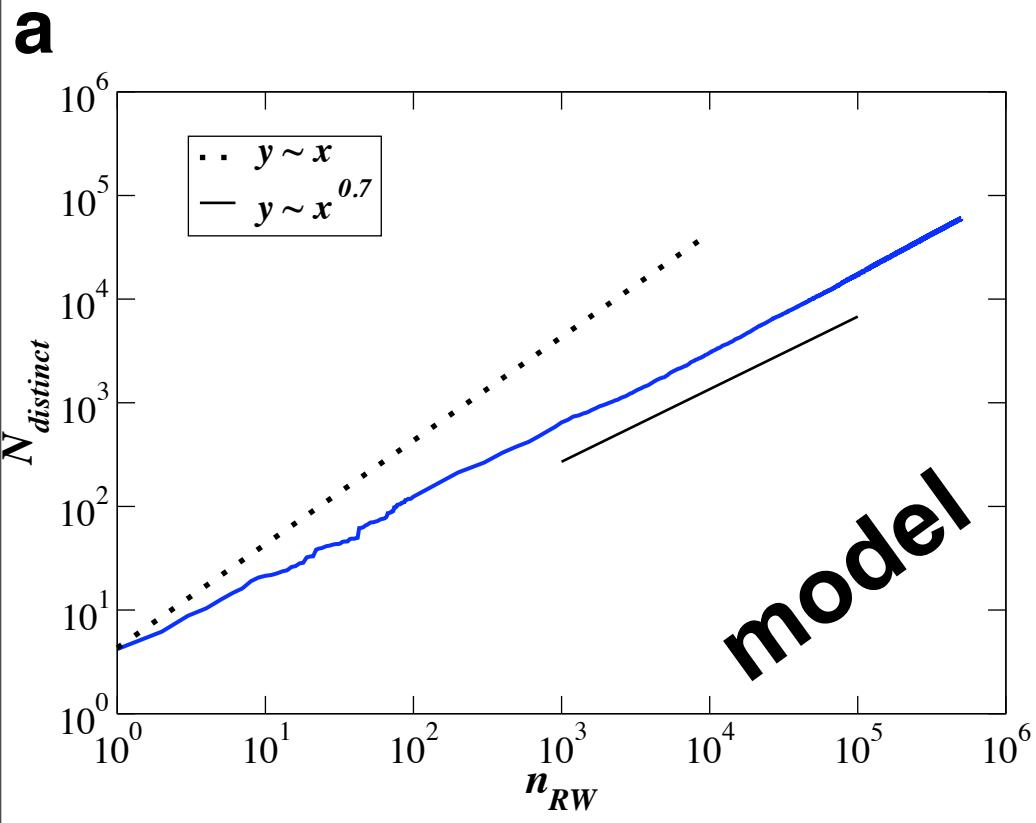
- tag frequencies
- dictionary size



Information Society
Technologies

Tagora comparison with real systems

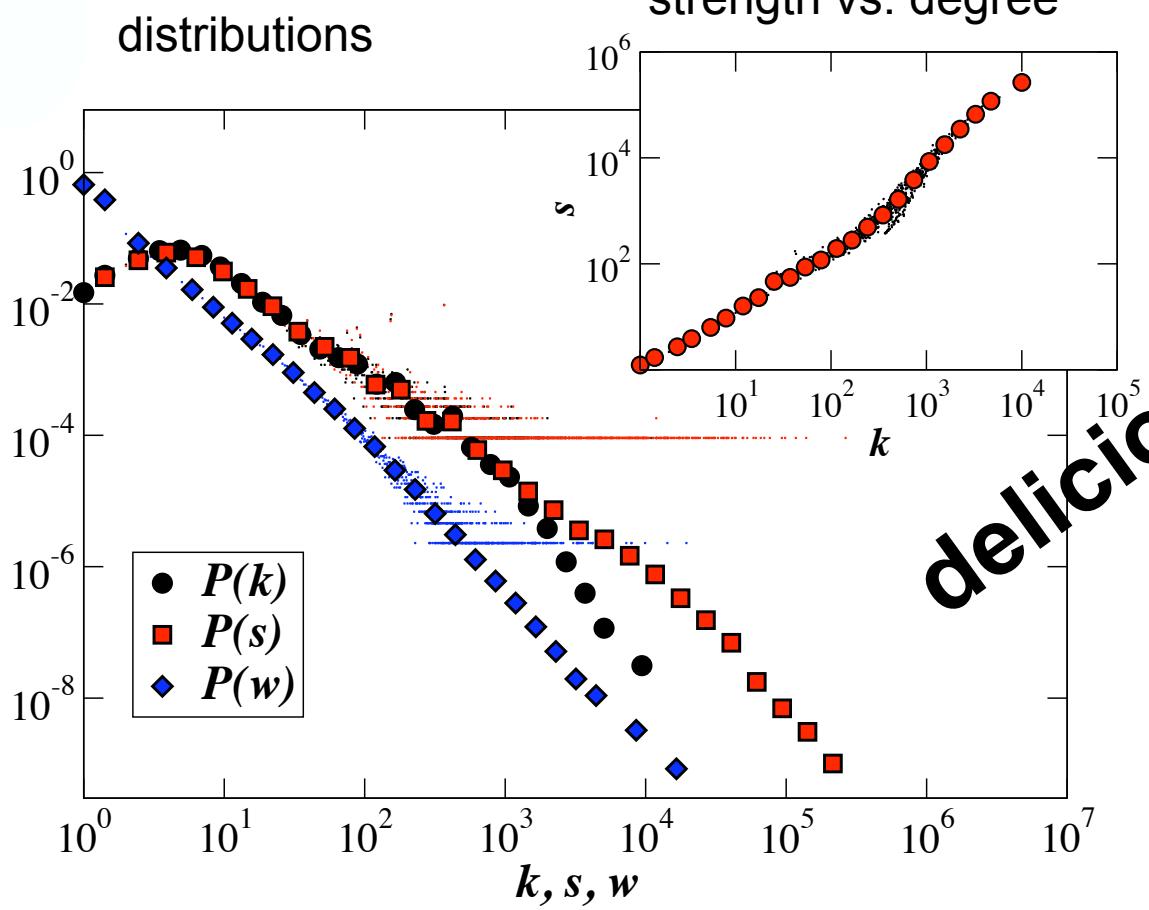
- tag frequencies
- dictionary size



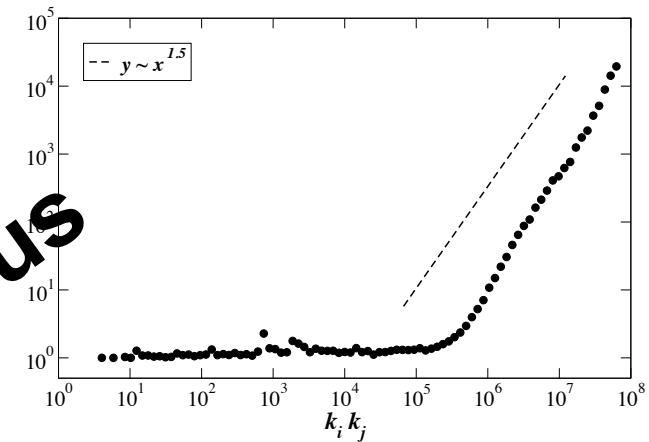
Information Society
Technologies

Tagora comparison with real systems

- single node: *degree, strength, edges weight*



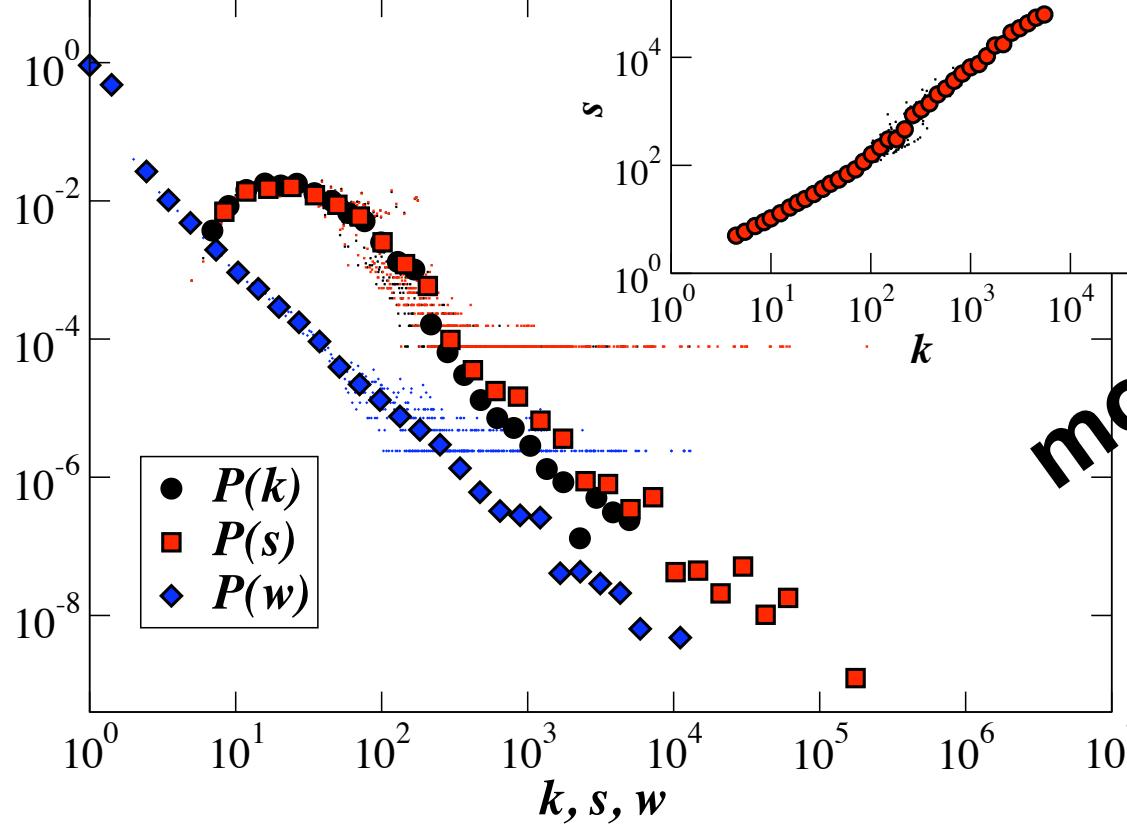
weight vs. degree



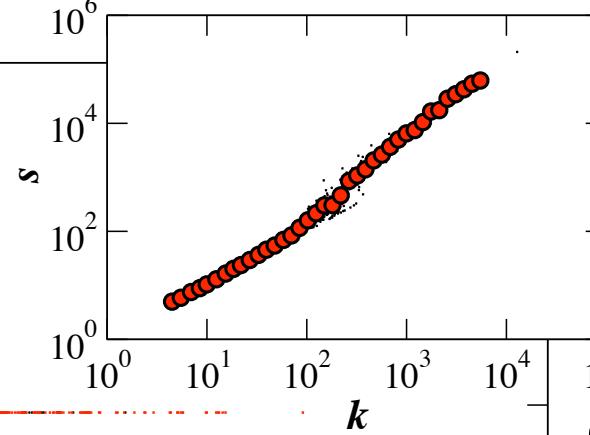
Tagora comparison with real systems

- single node: *degree, strength, edges weight*

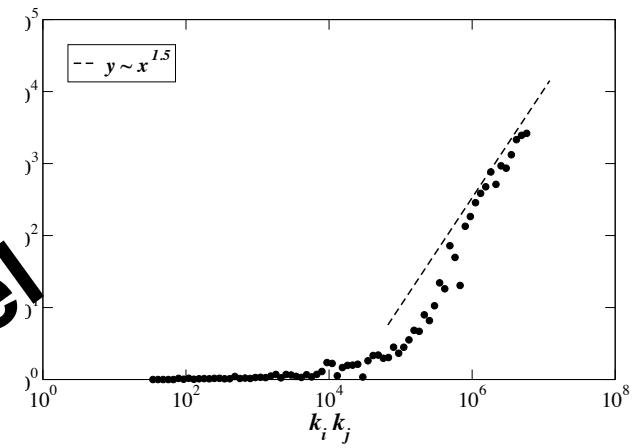
distributions



strength vs. degree



weight vs. degree



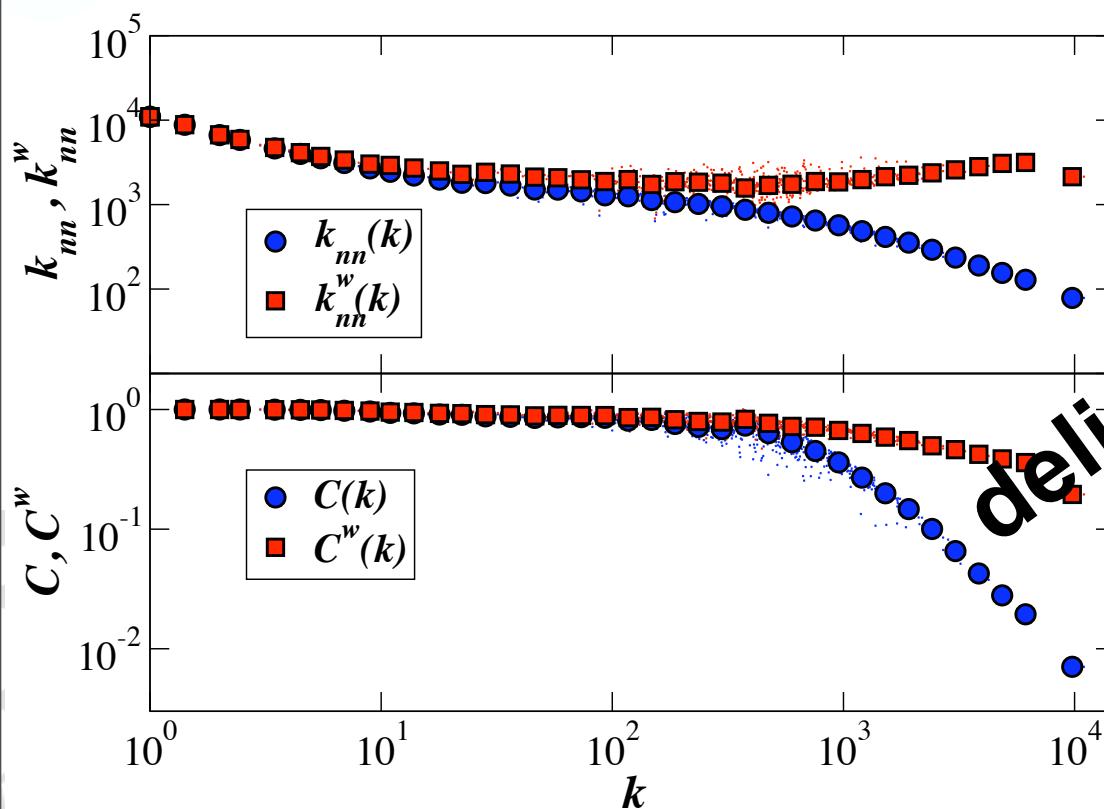
model



Information Society
Technologies

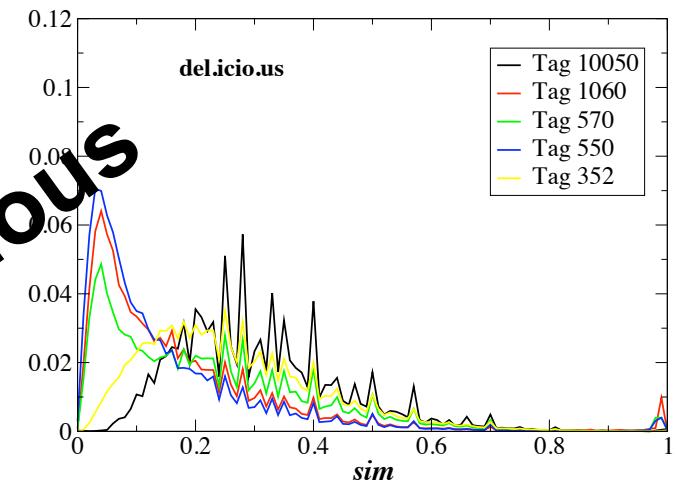
Tagora comparison with real systems

- Average nearest neighbor degree
- Clustering coefficients



delicious

Distribution of pair-wise structural similarities

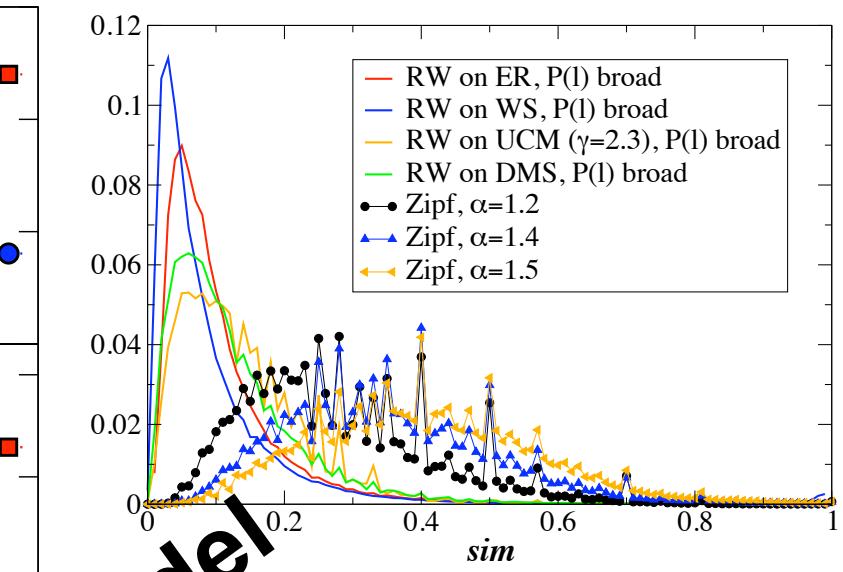
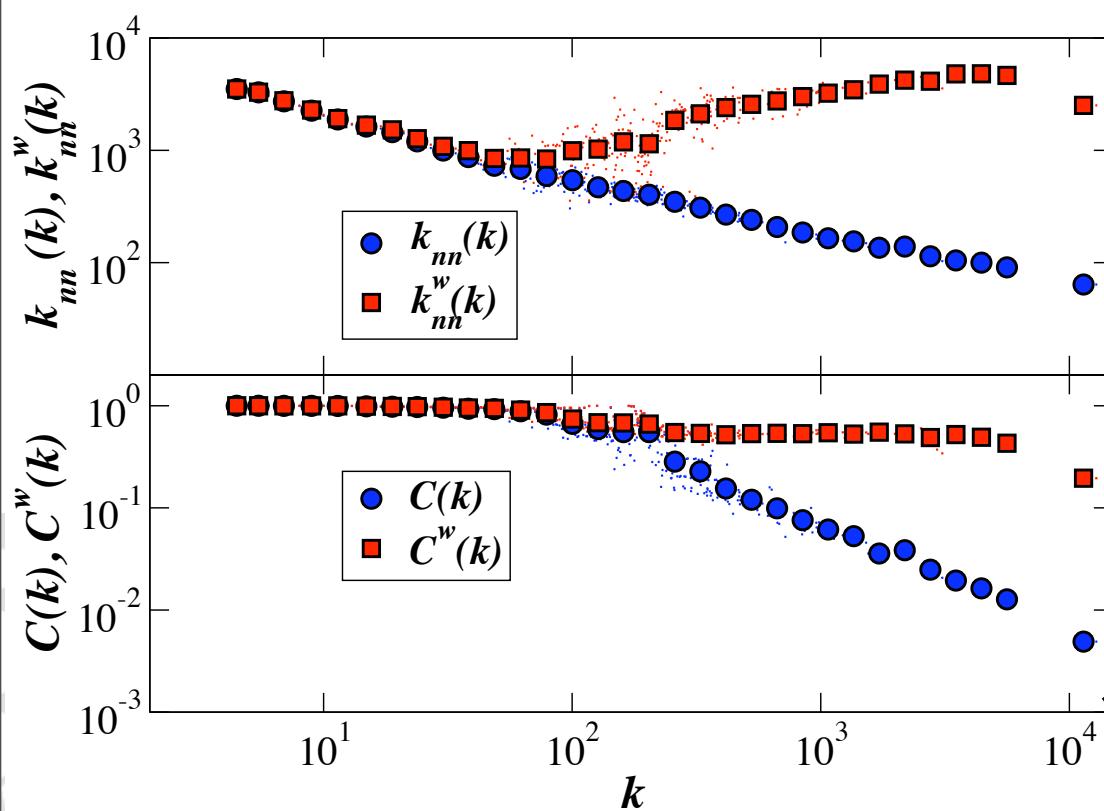


Information Society
Technologies

Tagora

comparison with real systems

- Average nearest neighbor degree
- Clustering coefficients



Information Society
Technologies

“Collective dynamics of social annotation”
C.Cattuto, A.Barrat, A.Baldassarri, G.Schehr, V.Loreto

PNAS, June, 2009