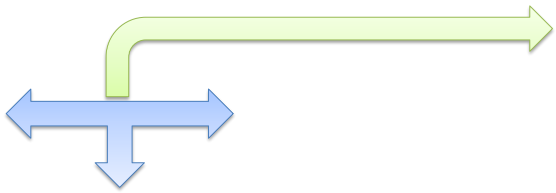
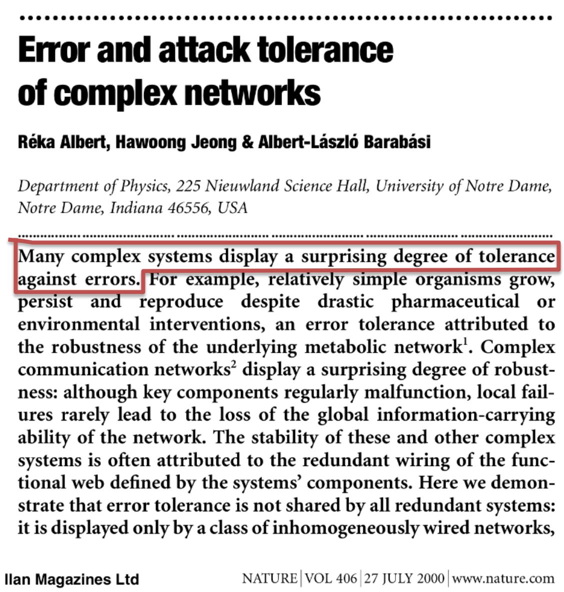
Complex **Directed** Networks: Dynamics & Communities



Hawoong Jeong (KAIST, Korea)

Y.D. Kim(KAIST, Korea→Samsung)
S.H. Lee(Umea Univ., Sweden→Oxford U., UK)
Y.H. Eom(ISI, Torino, Italy → U. Toulouse, France)
S.W. Son(U. Calgary, Canada→Hanyang U., Korea)
P.-J. Kim (UIUC→APCTP, Korea),
Y.Y. Ahn (NEU→Indiana U., USA)

Internet is fragile? (for our defense)



Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA

Many complex systems display a surprising degree of tolerance against errors. For example, relatively simple organisms grow, persist and reproduce despite drastic pharmaceutical or environmental interventions, an error tolerance attributed to the robustness of the underlying metabolic network[1]. Complex communication networks[2] display a surprising degree of robustness: although key components regularly malfunction, local failures rarely lead to the loss of the global information-carrying ability of the network. The stability of these and other complex systems is often attributed to the redundant wiring of the functional web defined by the systems' components. Here we demonstrate that error tolerance is not shared by all redundant systems: it is displayed only by a class of inhomogeneously wired networks,

llan Magazines Ltd

NATURE | VOL 406 | 27 JULY 2000 | www.nature.com

Please blame the Editor of the Nature ;) not us!

Why directed network? Interestingly,

Network Science

- After 2 key papers, D. Watts, S. Strogatz (1998) and A.-L. Barabasi, R. Albert (1999)

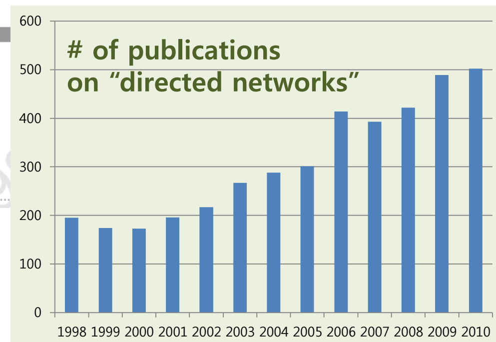
Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert



First empirical measurement on scale-free network was WWW, directed networks! *Nature* (1999)

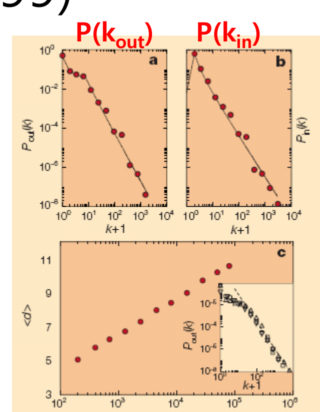
Internet

Diameter of the World-Wide Web

Réka Albert, Hawoong Jeong, Albert-László Barabási

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

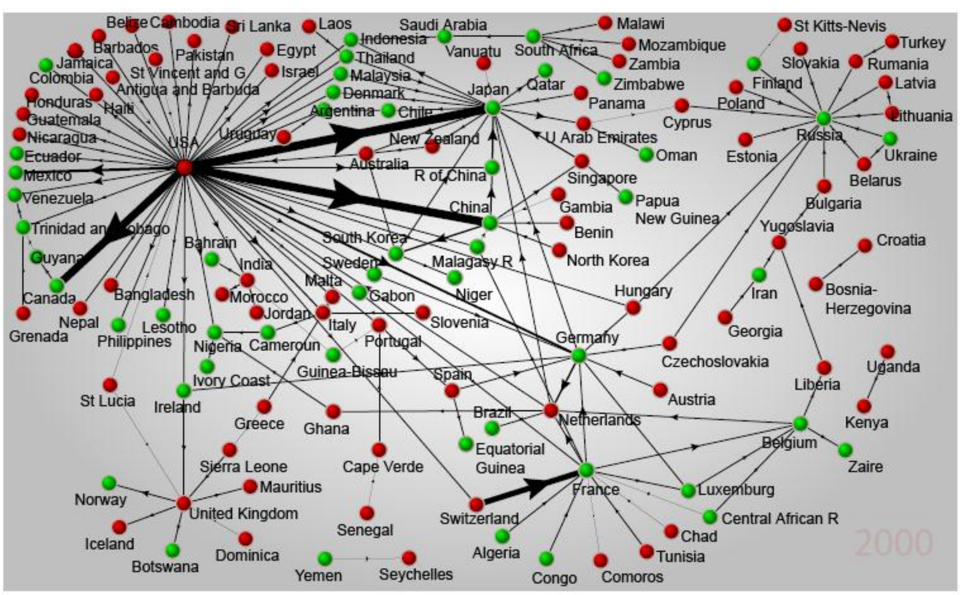
e-mail:alb@nd.edu



Many Directed Networks

- Communication Networks: World Wide Web
- Economic Networks: World trade web…

nodes: country links: import/export



Backbone of the world trade system.

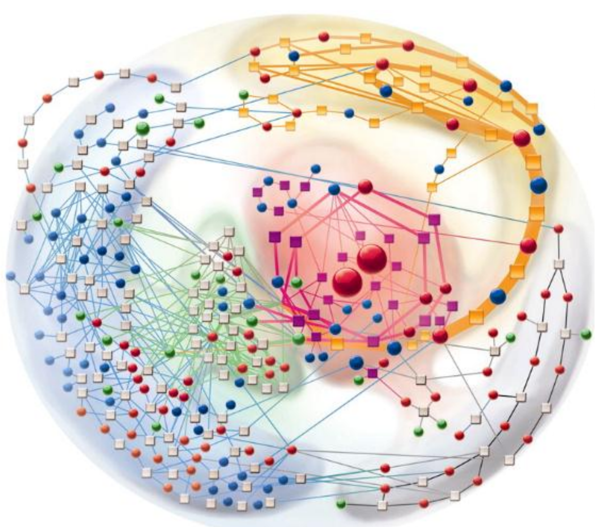
Serrano et al (2007)

Many Directed Networks

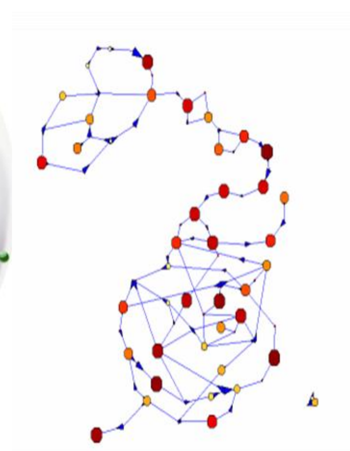
- Communication Networks: World Wide Web
- Economic Networks: World trade web…
- Biological Networks: Metabolic network, neural network, cortical network, gene regulatory network, food web …

nodes: metabolites　　links: bio/chemical reactions

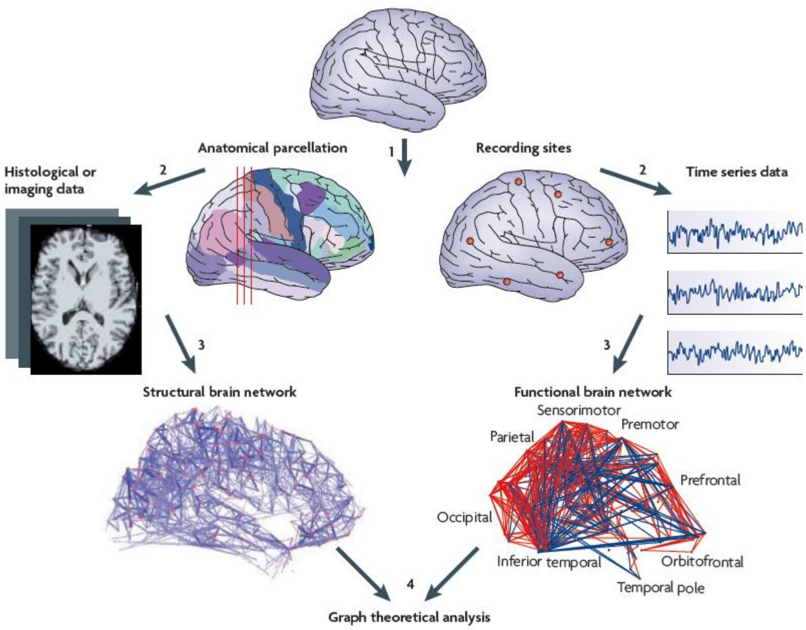
nodes: neurons　　links: synapses/correlation



E. coli metabolism



P.J. Kim et al (2007)



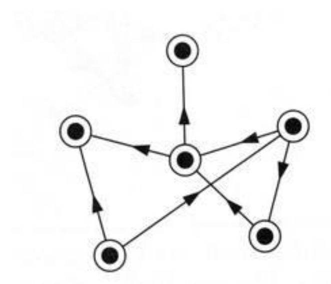
Structural and functional brain networks http://www.nature.com/nrn/journal/v10/n3/pdf/nrn2575.pdf

Many Directed Networks

- Communication Networks: World Wide Web
- Economic Networks: World trade web…
- Biological Networks: Metabolic network, neural network, cortical network, gene regulatory network, food web …
- Social Networks: Friendship network, email network, phone call network etc

  nodes: people   links: social relationship

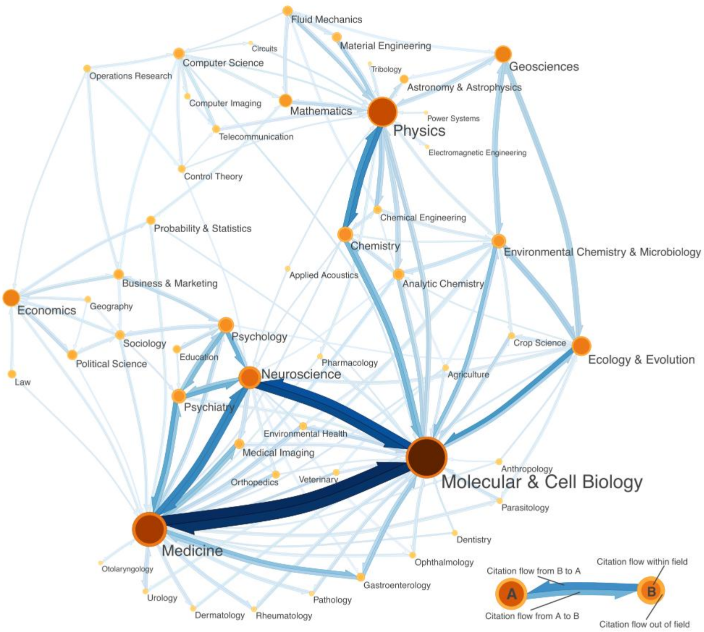
Facebook vs twitter



Many Directed Networks

- Communication Networks: World Wide Web
- Economic Networks: World trade web…
- Biological Networks: Neural network, cortical network, metabolic network, gene regulatory network, cell cycle network, food web …
- Social Networks: Friendship network, email network, phone call network etc
- Other networks: Citation network, Word network etc

nodes: papers   links: references/citations

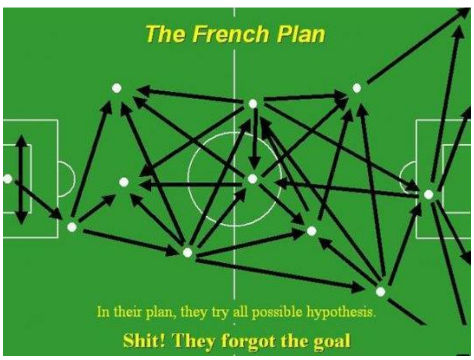


A map of science based on citation patterns. Rosvall et al (2007)

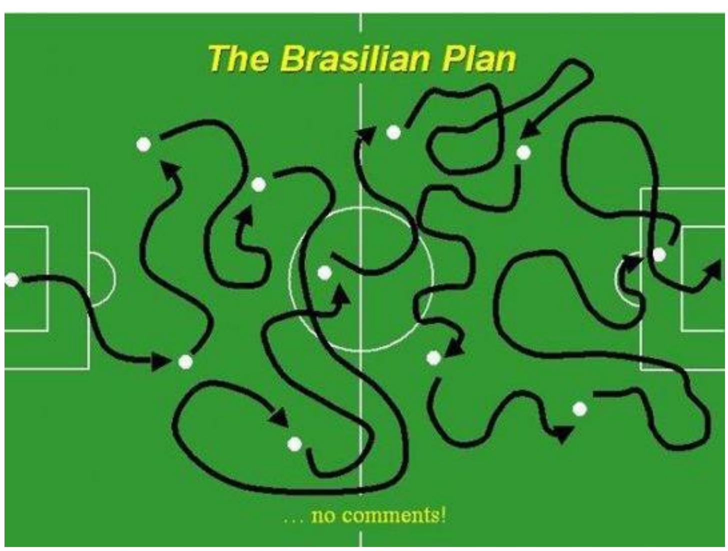
There are few funny cartoons I found from the Internet, which is directed networks… (Don't blame me, I didn't draw!)


The English Plan
Depending on the wind, the striker's position may vary…


The German Plan
Radical, efficient, unstoppable… (ball's speed may reach 297 km/h)

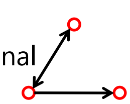

The French Plan
In their plan, they try all possible hypothesis.
Shit! They forgot the goal


The Italian Plan
Iron defense, small ideas in midfield, passes to striker..and…Penalty


The Brasilian Plan
… no comments!

Basic Properties of Directed Networks

Reciprocity, r



Q: Do they know each other?
Or just one way connection?

$$r \equiv \frac{L^{\leftrightarrow}}{L} = \frac{\text{\# of links pointing in both directions}}{\text{total \# of links}}$$

- Tells how much information would be lost if the direction is ignored.
- Can be used as a criterion to verify a model.
- Related to propagation process on the network
- Influences degree dist. & degree correlation

Modified reciprocity

Limitations of original def.:

1. not compared with random network
2. link density affects reciprocity:
   high link density -> high reciprocity

$$\rho \equiv \frac{\sum_{i \neq j}(a_{ij} - \bar{a})(a_{ji} - \bar{a})}{\sum_{i \neq j}(a_{ij} - \bar{a})^2}, \quad = \frac{L^{\leftrightarrow}/L - \bar{a}}{1 - \bar{a}} = \frac{r - \bar{a}}{1 - \bar{a}}.$$

$$a \equiv \frac{L}{N(N-1)}$$

the ratio of observed to possible directed links

Garlaschelli (2004)

+

Network	ρ	σ_ρ	$\rho_{\min}$
Perfectly reciprocal	1	...	$-\frac{\bar{a}}{1-\bar{a}}$
World Trade Web (53 webs) [10]			
Most correlated (year 2000)	0.952	0.002	($\bar{a} > 0.5$)
Least correlated (year 1948)	0.68	0.01	−0.80
World Wide Web [7]	0.5165	0.0006	−0.0001
Neural networks [13,14]			
Neuron classes	0.44	0.03	−0.04
Neurons	0.41	0.02	−0.03
Email networks [5,6]			
Address books	0.231	0.003	−0.001
Actual messages	0.194	0.002	−0.001
Word networks [15]			
Dictionary terms	0.194	0.005	−0.002
Free associations	0.123	0.001	−0.001
Cellular networks (43 webs) [16]			
Most correlated (*H. influenzae*)	0.052	0.006	−0.001
Least correlated (*A. thaliana*)	0.006	0.004	−0.003
Areciprocal	0	...	$-\frac{\bar{a}}{1-\bar{a}}$
Shareholding networks [17]			
NYSE	−0.0012	0.0001	−0.0012
NASDAQ	−0.0034	0.0002	−0.0034
Food webs [11,12]			
Silwood Park	−0.0159	0.0008	−0.0159
Grassland	−0.018	0.002	−0.018
Ythan Estuary	−0.031	0.005	−0.034
Little Rock Lake	−0.044	0.007	−0.080
Adirondack lakes (22 webs)			
Most correlated (B. Hope)	−0.06	0.02	−0.10
Least correlated (L. Rainbow)	−0.102	0.007	−0.102
St. Marks Seagrass	−0.105	0.008	−0.105
St. Martin Island	−0.13	0.01	−0.13
Perfectly antireciprocal	−1	...	−1

−

Prey <-> Predator ??

Garlaschelli (2004)

Single node properties:

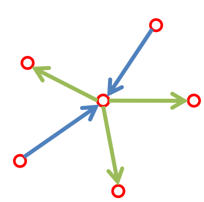
Centrality : which node is important?

"Directed" Degree

in-degree $k_i^{in} \equiv \sum_j A_{ji}$

out-degree $k_i^{out} \equiv \sum_j A_{ij}$

degree $k_i \equiv k_i^{in} + k_i^{out}$



degree distribution $P(k^{in})$ $P(k^{out})$ $P(k)$

In many real-world networks, in- & out-degree have different meaning and different generating mechanism

In World Wide Web,

In-degree & out-degree dist. show different behavior (different exponent)
High in-degree: authority; high out-degree: hub (portal)

Network	N	<k>	γ_{in}	γ_{out}	L	Ref.
WWW (Web page)	325,729	4.51	2.1	2.45	11.2	Albert1999
WWW (Web page)	4×10^7	7	2.1	2.38		Kumar1999
WWW (Web page)	2×10^8	7.5	2.1	2.72	16	Broder2000

The webpage owner can only decide the out-degree, not in-degree(fame).

You can't increase in-degree by yourself... **BUT**

In Wikipedia

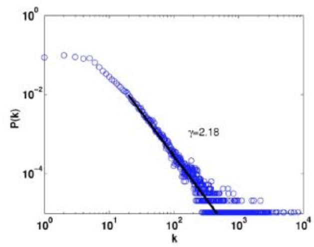


FIG. 3. (Color online) The probability distribution of the in-degree for the Japanese Wikipedia.

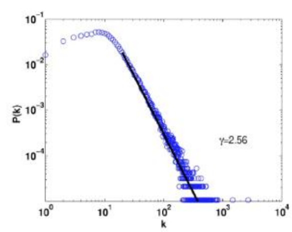
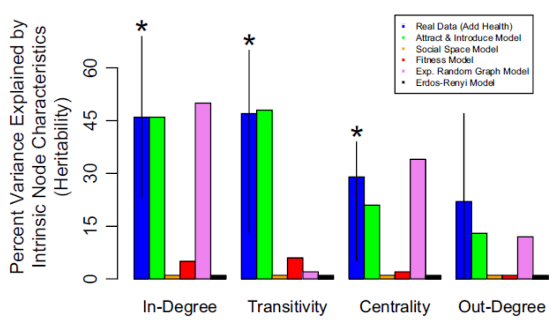


FIG. 4. (Color online) The probability distribution of the out-degree for the Japanese Wikipedia.

Zlatic2006

In human social network,

in-degree (how many times a person is named as a friend), ~ popularity
out-degree (how many friends a person names) ~ social activity



Fowler2009

A research based on twin study shows **in-degree**, transitivity(clustering coefficient), centrality are found to be **heritable.**

If your parents are famous, you have a chance to become famous too!

"Directed" Centrality (closeness, betweenness)

Straightforward in the SCC[giant cluster]

$$c_{Cl}(v) \equiv \frac{1}{\sum_{u \in V} dist(v,u)}$$

$$c_B(v) \equiv \sum_{s \neq t \neq v \in V} \frac{\sigma(s,t\,|\,v)}{\sigma(s,t)}$$

$\sigma(s,t\,|\,v)$ is the total number of shortest paths between s and t that pass through v

PageRank®

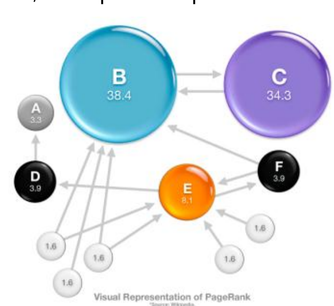
is the probability that the random walker visiting node *i* at the stationary state.

- A page with more recommendations (more incoming links) is considered to be more important.
- A webpage is considered important if it is pointed to by other important pages.

Undirected networks: PageRank ~ k , local information Directed networks: PageRank is related to global structure.

- Rank web pages (Google)
- academic papers (eigenfactor.org), eigenfactor, to replace impact factor

Related quantities (spectrum based):
Eigenvector Centrality (Bonacich1972)
Influence
Katz status index (Katz1953)



Used for Directed network community

- **Google Matrix** $G_{ij} = \alpha H_{ij} + \frac{1}{n}(\alpha a_i + 1 - \alpha)$

  Where $H_{ij} = A_{ij} / k_i^{out}$, and $a_i = 1$ if and only if i is a dangling node.

  - G_{ij} is probability that the random walker moves from node *i* to *j* at next step when it is visiting node *i*.
  - α and a_i is added to avoid random walker being trapped in dangling nodes and trap region.
  - G is a completely dense, stochastic, and primitive matrix. There always exists the stationary vector π^T.

- **PageRank** $\pi^T G = \pi^T$

  - π_i is the probability that the random walker visiting node *i* at the stationary state.

PageRank Distribution of *.brown.edu website

Pandurangan2002

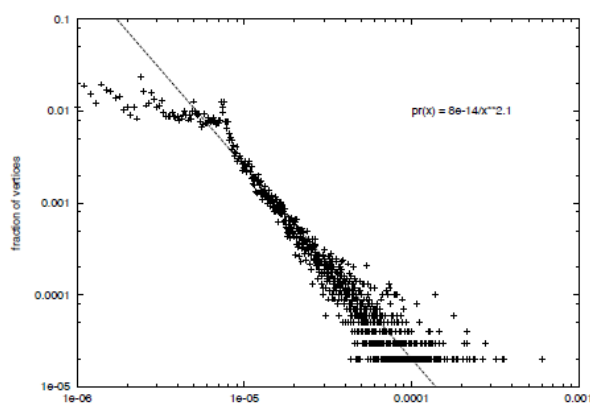


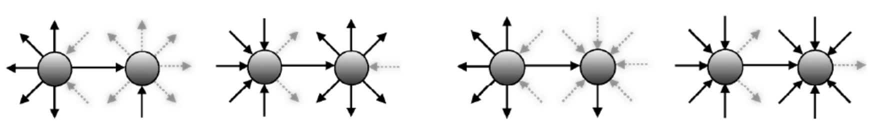
Figure 3. Log-log plot of the PageRank distribution of the Brown domain (*.brown.edu). A vast majority of the pages (except those with very low PageRank) follow a power law with exponent close to 2.1. The plot almost flattens out for pages with very low PageRank.

More than single node:

2-node correlations & more…

"directed" Assortativity

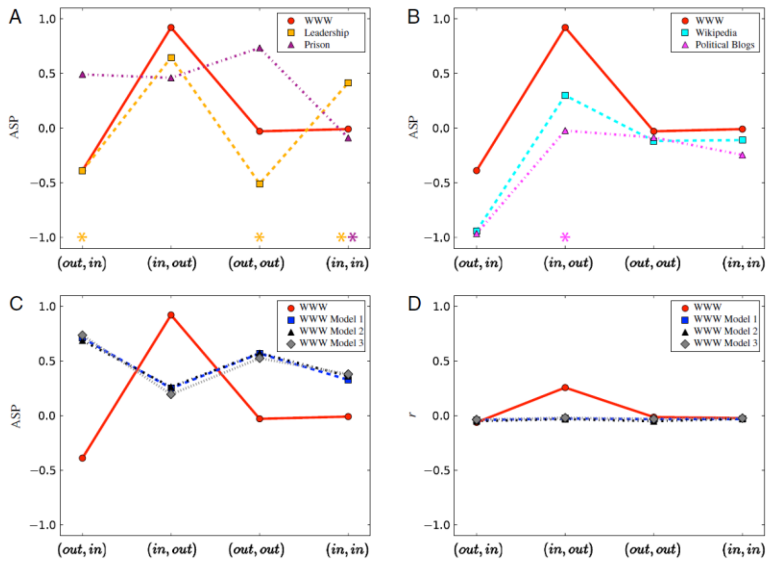
Foster (2009)



$$r(\alpha,\beta) = \frac{E^{-1}\sum_i [(j_i^\alpha - \overline{j^\alpha})(k_i^\beta - \overline{k^\beta})]}{\sigma^\alpha \sigma^\beta}$$

$$Z(\alpha,\beta) = \frac{r_{\mathrm{rw}}(\alpha,\beta) - \langle r_{\mathrm{rand}}(\alpha,\beta)\rangle}{\sigma[r_{\mathrm{rand}}(\alpha,\beta)]}.$$

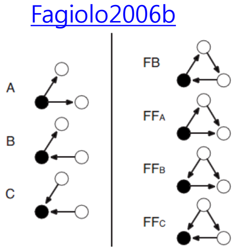
$$\mathrm{ASP}(\alpha,\beta) = \mathrm{Z}(\alpha,\beta)/(\textstyle\sum_{\alpha,\beta} \mathrm{Z}(\alpha,\beta)^2)^{1/2}.$$



- WWW and social network show different profile.
- WWW, wikipedia and political blogs are also slightly different.
- Assortativity can be used to check the validity of models.

"directed" Clustering coefficient

Fagiolo2006b



$$\mathbf{C}^{(i)} = \left(\frac{N^{(i)}_{\mathrm{FB}}}{M^{(i)}_B}, \frac{N^{(i)}_{\mathrm{FF}_A}}{M^{(i)}_A}, \frac{N^{(i)}_{\mathrm{FF}_B}}{M^{(i)}_B}, \frac{N^{(i)}_{\mathrm{FF}_C}}{M^{(i)}_C} \right),$$

where

$$M^{(i)}_A = \sum_{j,k} a_{ij} a_{ik}, \quad M^{(i)}_B = \sum_{j,k} a_{ij} a_{ki}, \quad M^{(i)}_C = \sum_{j,k} a_{ji} a_{ki},$$

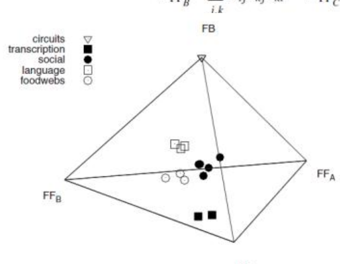
$$N^{(i)}_{\mathrm{FB}} = \sum_{j,k} a_{ij} a_{jk} a_{ki}, \quad N^{(i)}_{\mathrm{FF}_A} = \sum_{j,k} a_{ij} a_{kj} a_{ik},$$

$$N^{(i)}_{\mathrm{FF}_B} = \sum_{i,k} a_{ij} a_{kj} a_{ki}, \quad N^{(i)}_{\mathrm{FF}_C} = \sum_{i,k} a_{ji} a_{kj} a_{ki},$$

$$\tilde{\mathbf{C}}^{(i)} = \frac{\mathbf{C}^{(i)}}{T^{(i)}},$$

$$T^{(i)} = \frac{N^{(i)}_{\mathrm{FB}}}{M^{(i)}_B} + \frac{N^{(i)}_{\mathrm{FF}_A}}{M^{(i)}_A} + \frac{N^{(i)}_{\mathrm{FF}_B}}{M^{(i)}_B} + \frac{N^{(i)}_{\mathrm{FF}_C}}{M^{(i)}_C}$$

Ahnert2008



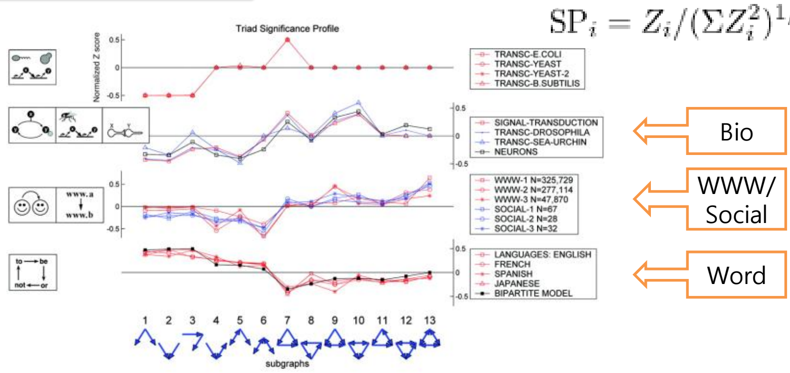
Networks of the same type are located in the same region of Clustering signature tetrahedron.

Motif : basic building block

13 possible triads.

$$Z_i = (Nreal_i - \langle Nrand_i \rangle)/\mathrm{std}(Nrand_i)$$

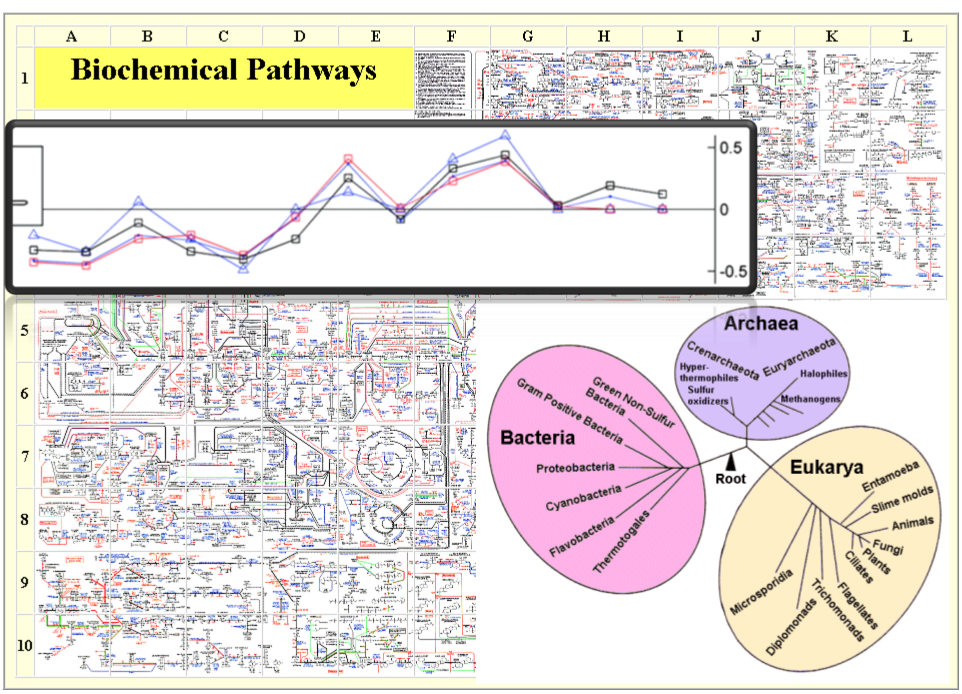
$$SP_i = Z_i/(\Sigma Z_i^2)^{1/2}$$



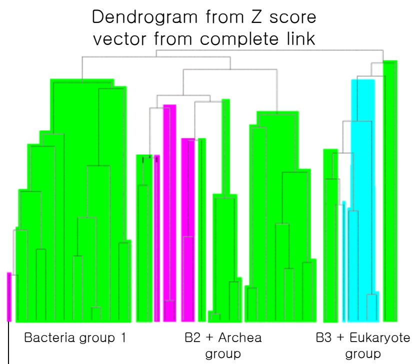
Motif profile **evolved** to perform similar tasks…

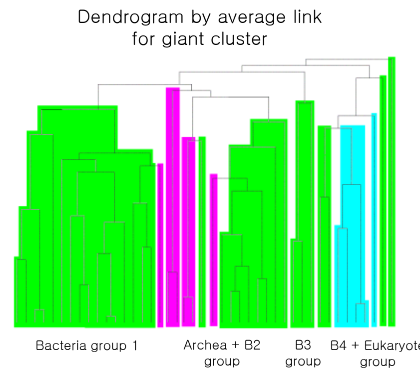
Applying to metabolic networks (Archea, Bacteria, Eukaryote) of organisms has **different kinds of "common motifs" in details**

➔ Can be used for clustering (taxnomy !) [Milo et al 2004]



Motif Clustering of 43 organisms





- Analysis
  - In large scale, **we identified 3 domains of life,** Archae, Bateria, and Eukaryote of 43 organisms .
  - One exceptional species of Archea can be explained because it belongs to different phylum as compared to other 5 species.

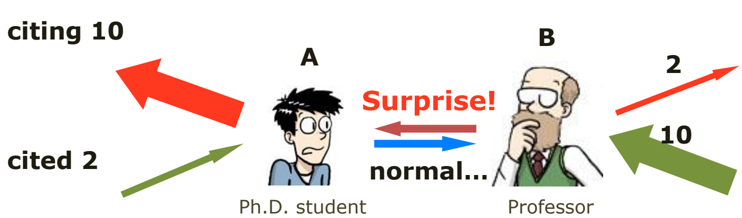
→ Aeropyrum pernix (name), Crenarchaeota (phylum : 동물 분류 상의 문(門)), thermophilic(열을 좋아하는 성질)

Y. Eom, S. Lee, H. Jeong (J. Theo. Bio. 2006)

What if we only have undirected network?

There is a simple way to construct directed network out of undirected (but weighted) network!

We start from the fact that all links between nodes are not equivalent!



S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)

Nature (2008/9): Google 10th anniversary

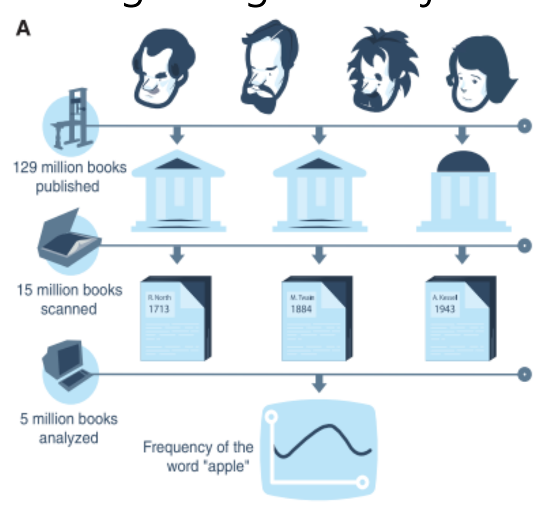
- **BIG DATA: PetaByte Era**
- **Data wrangling**
- **Welcome to the petacentre**

Let's see what google & we can do!

- **Distilling meaning from data**
- **The Harvard computers**
- **The future of biocuration**

Last year in Google…

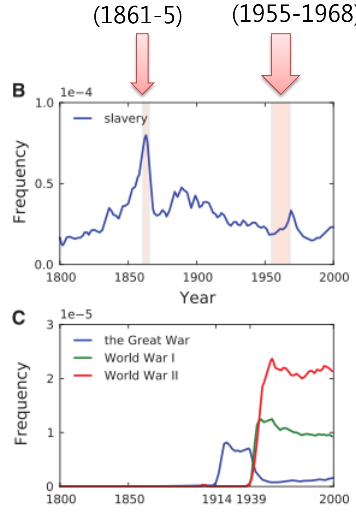
- Google N-gram Project!



J.-B. Michael et al Science (2011)

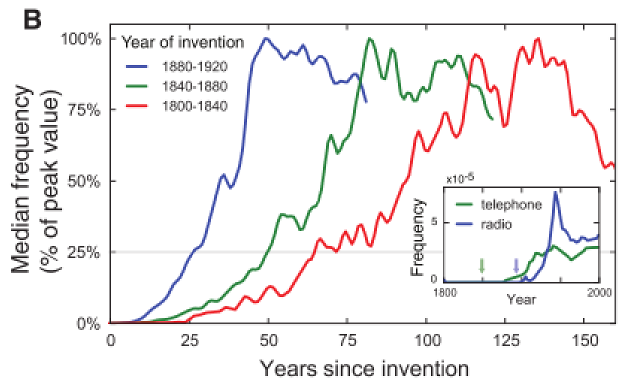
Civil War (1861-5)

Civil right movements (1955-1968)



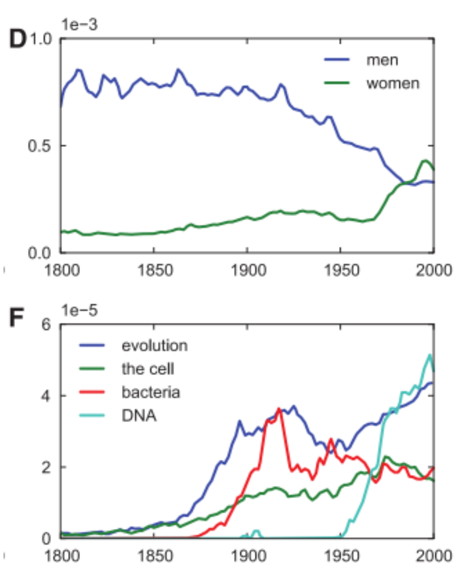
You can do many things with this N-gram data

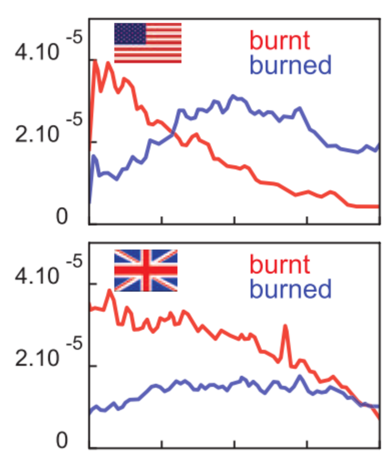
- Culturomics : Finding culture trends! E.g. cultural adoption is getting quicker!



J.-B. Michael et al Science (2011)

Quantitative analysis possible in history/science …

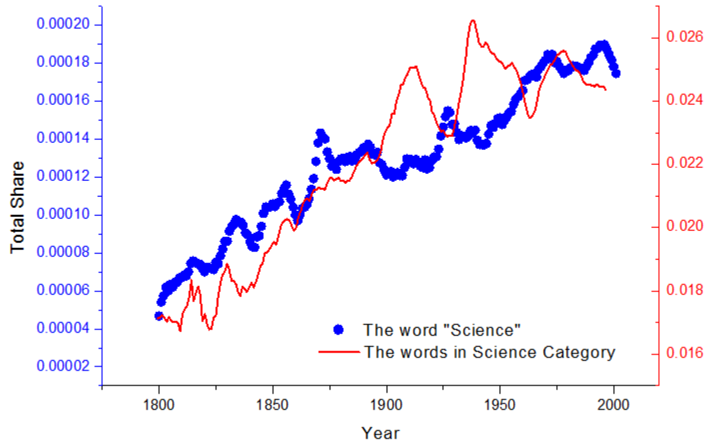




J.-B. Michael et al Science (2011)

Which technology wins the race?

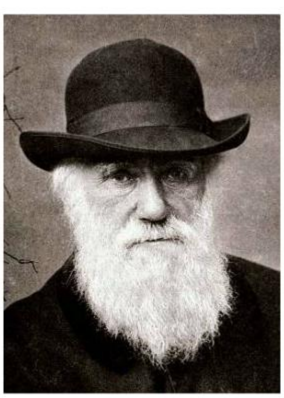
Science vs Science



J. Yun, P.-J. Kim, H. Jeong (in preparation)

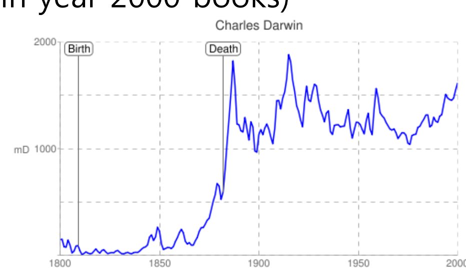
Science Hall of Fame

- Looking up the names of Scientists
  (collected from Wikipedia ~ 7,500 → 4,169 after filtering)



Charles Darwin

Charles Darwin : 148,429 times appeared
(2% of English books) & increasing
(4% in year 2000 books)



UNIT: if average annual frequency=Darwin
→ 1 Darwin unit
1 Darwin = 1000 mili-Dawin (mD)

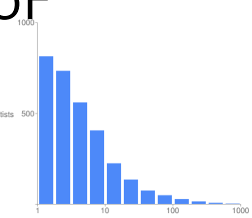
A. Veres et al Science (2011)

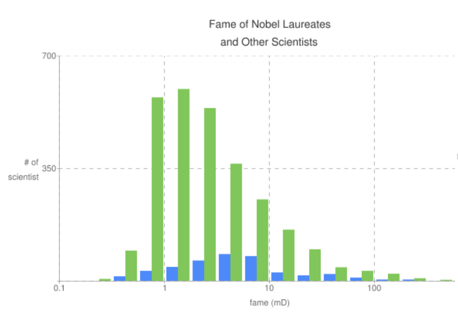
Science Hall of Fame (N-gram)

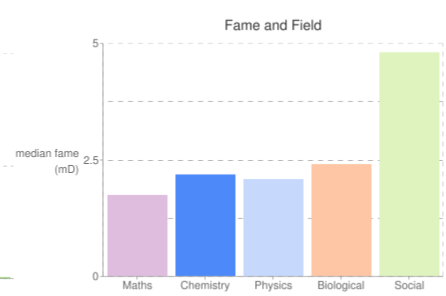
Full Name	Born	Died	milliDarwins	Nobel Prize Winner		
				Bio	Chem	Phy
Bertrand Russell	1872	1970	1500			
Charles Darwin	1809	1882	1000			
Albert Einstein	1879	1955	878			✓
Lewis Carroll	1832	1898	479			
Claude Bernard	1813	1878	429			
Oliver Lodge	1851	1940	394			
Julian Huxley	1887	1975	350			
Karl Pearson	1857	1936	346			
Niels Bohr	1885	1962	289			✓
Alexander Graham Bell	1847	1922	274			
Max Planck	1858	1947	256			✓
Francis Galton	1822	1911	255			
Robert Oppenheimer	1904	1967	252			
Louis Pasteur	1822	1895	237			
Chaim Weizmann	1874	1952	236			
Alfred North Whitehead	1861	1947	229			
Marie Curie	1867	1934	189		✓	✓
Robert Koch	1843	1910	185	✓		
Isaac Asimov	1920	1992	183			

Few things on SHoF

- Distribution of mD ~ power-law
- Not all Nobel winners are at top (80% are less than 10mD)





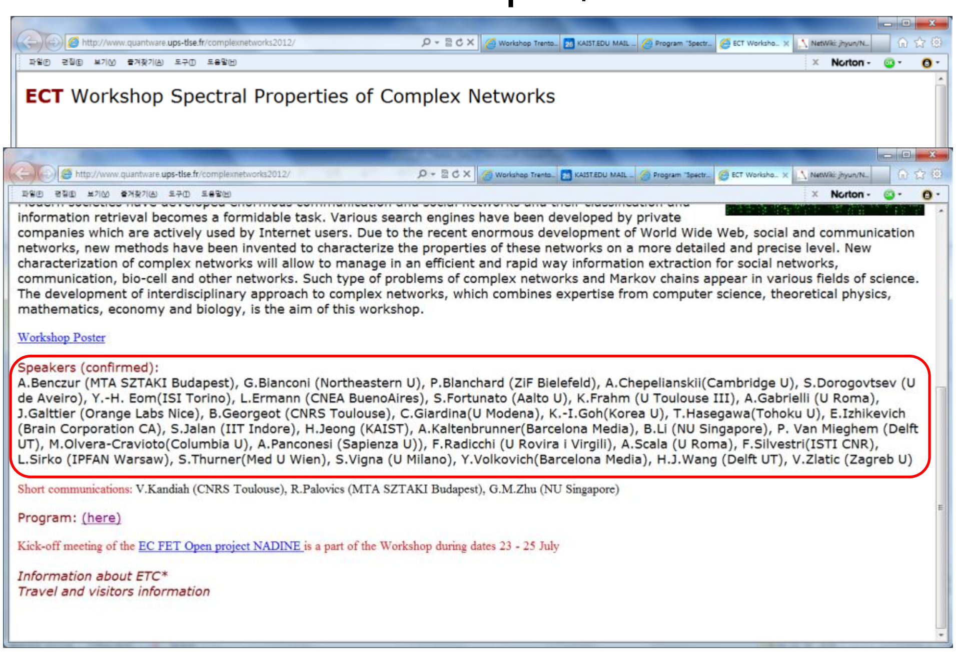


A. Veres et al Science (2011)

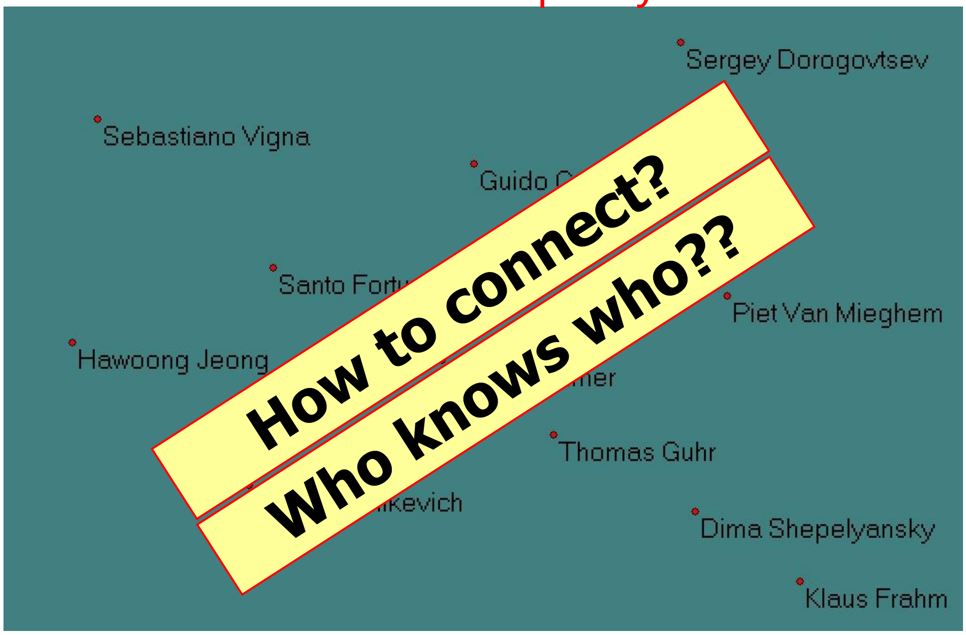
Any tips for immortal fame?

- Seek the social sciences: avoid mathematics
  - The most famous living scientist, Noam Chomsky 507mD
  - Other living scientist, Barry Commoner 109mD, @Queens College in NY, 93 years old, pioneer of ecology & the environmental movement
  - in Math, top = Bertrand Russel, 1500mD, author of Principia Mathematica) but "He's famous for not doing math!" by Michael Thaddeus @Columbia (0mD) & Mathematician doesn't care much…
- Do good work, but don't get caught up in the citation rat race
  - One of the most highly cited scientists, Edward Witten has only 8mD
  - Paul Erdos (wrote 1400 papers) has 3.5mD only
- Write a popular book
  - Isaac Asimov (183 mD), Carl Sagan (152 mD), Rachel Carson (12mD), Richard Dawkins (90mD)
- Embrace controversy
  - Timothy Leary (136mD) research on psychedelic drugs (not to mention his use and advocacy, Richard Nixon called him "the most dangerous man in USA")
  - Richard Feynman (47mD) gets more fame after TV appearance (hearing on 1986 Space shuttle Challenger disaster)
  - Charles Darwin (1000 mD) himself gets fame from world-shaking controversy…

Back to our topic, network!

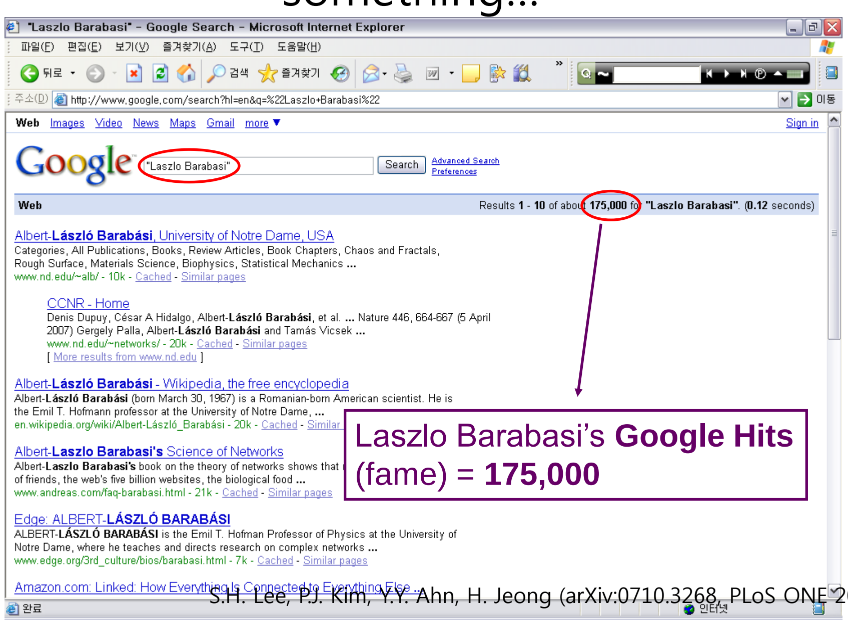


Trento Workshop Key talks



S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

Basic idea: Using search engine for finding something…



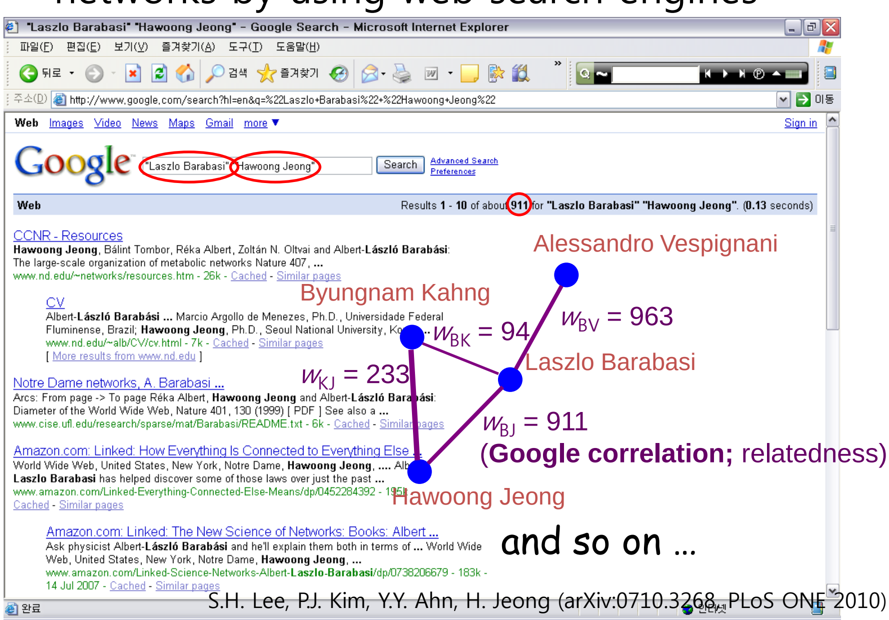
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

To make a network, we need "link" information between 2 persons… HOW?

Ask Google!

S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

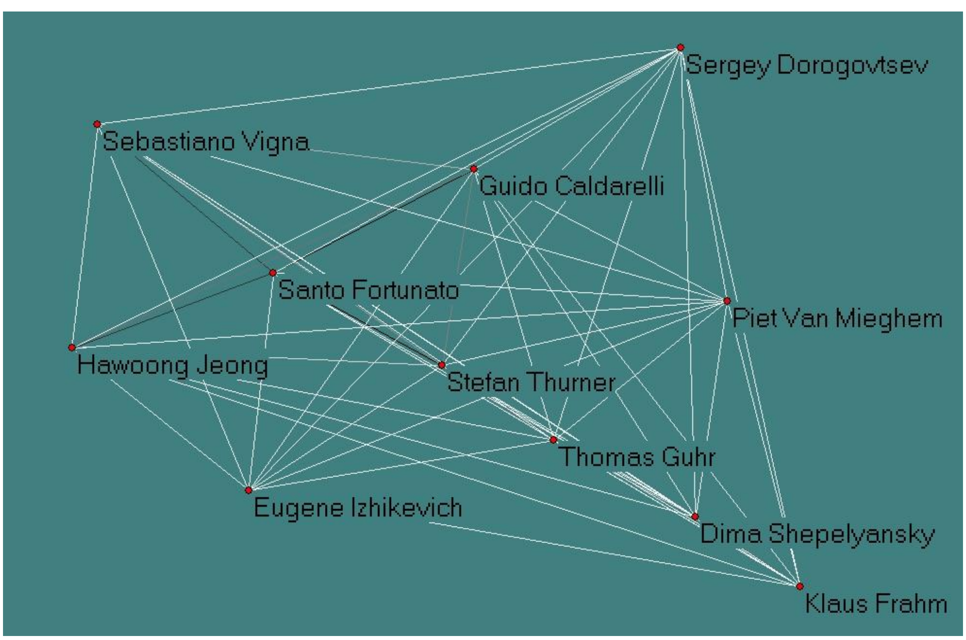
Basic idea: Constructing weighted social networks by using web search engines



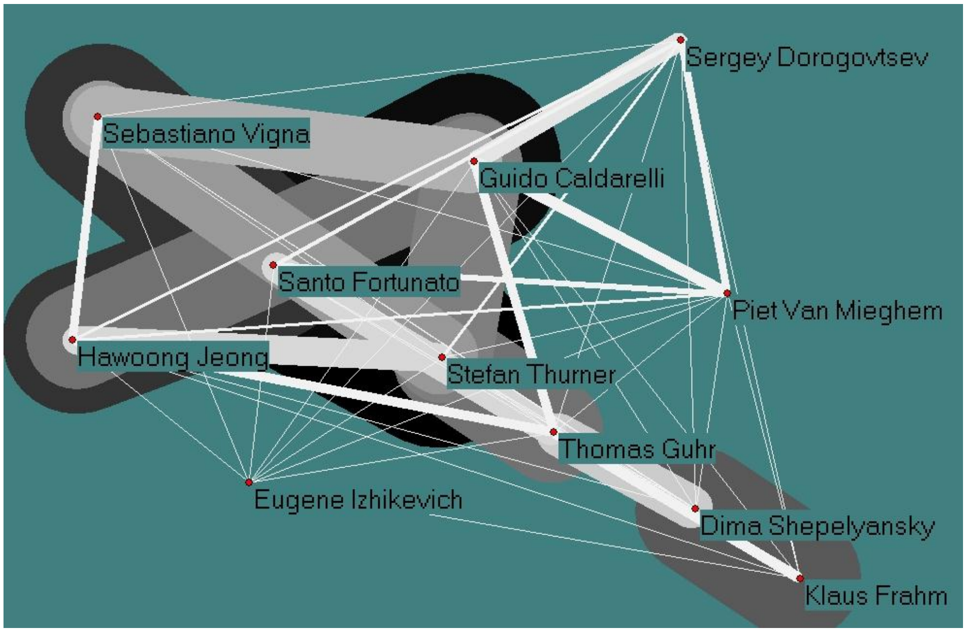
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

Stefan Thurner	&	Santo Fortunato	: 8960
Guido Caldarelli	&	Santo Fortunato	: 8800
Sebastiano Vigna	&	Santo Fortunato	: 7450
Santo Fortunato	&	Hawoong Jeong	: 7270
Dima Shepelyansky	&	Klaus Frahm	: 6180
Stefan Thurner	&	Thomas Guhr	: 5070
Guido Caldarelli	&	Hawoong Jeong	: 4970
Guido Caldarelli	&	Stefan Thurner	: 4830
Sebastiano Vigna	&	Stefan Thurner	: 3760
Guido Caldarelli	&	Sebastiano Vigna	: 3060
Dima Shepelyansky	&	Thomas Guhr	: 2080
Santo Fortunato	&	Thomas Guhr	: 1470
Stefan Thurner	&	Hawoong Jeong	: 1430

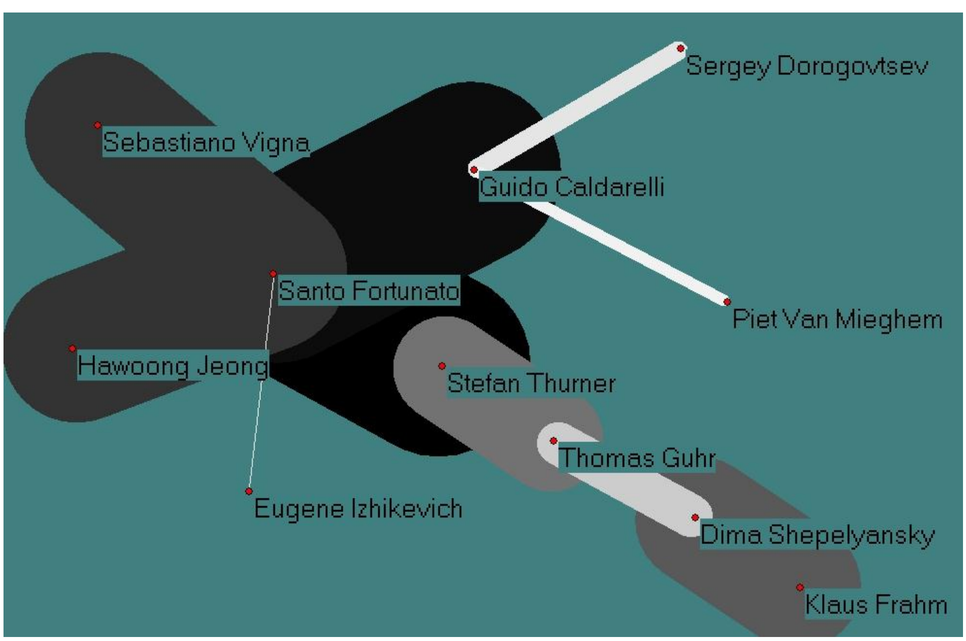
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)



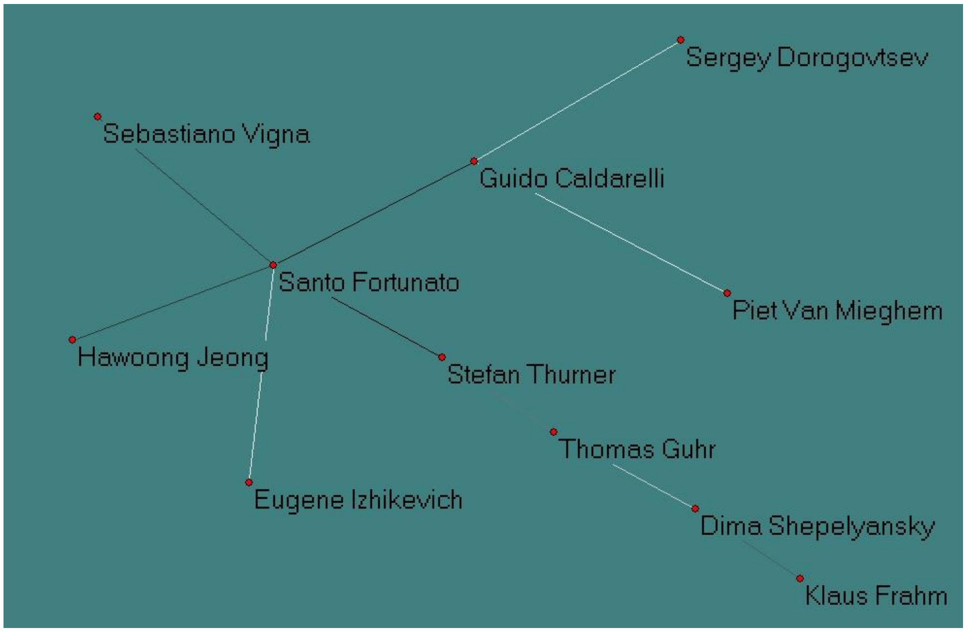
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)



S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)



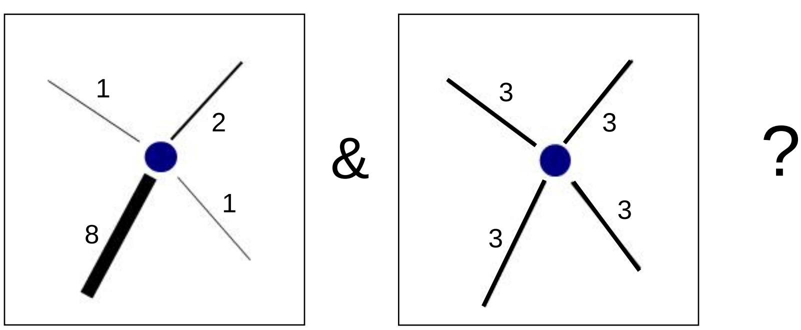
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)



S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

Let's consider weighted "undirected" network.

How to quantify the difference between …



Degree k = 4,
Strength s = 12

Degree k = 4,
Strength s = 12

(Strength s = Sum of its weights)

S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)

"Quantifying broad distribution of weights"

$$D_\alpha(\{\tilde{w}\}) = \left(\sum_j \tilde{w}_{ij}^\alpha\right)^{\frac{1}{1-\alpha}} \quad \text{where} \quad \tilde{w}_{ij} = \frac{w_{ij}}{s_i} = \frac{w_{ij}}{\sum_k w_{ik}}$$

$$\longrightarrow \quad D_{\alpha=0}(\{\tilde{w}\}) = \text{degree } k$$

$$\longrightarrow \quad \lim_{\alpha\to 1} D_\alpha(\{\tilde{w}\}) = \exp(-\sum_j \tilde{w}_{ij} \log \tilde{w}_{ij}) = \prod_j \tilde{w}_{ij}^{-\tilde{w}_{ij}}$$

$$\longrightarrow \quad D_{\alpha=2}(\{\tilde{w}\}) = \frac{1}{\sum_j \tilde{w}_{ij}^2}$$

$$\text{cf) } \log D_\alpha(\{\tilde{w}\}) = \frac{1}{1-\alpha} \log\left(\sum_i \tilde{w}_{ij}^\alpha\right) \equiv \text{Rényi entropy}$$

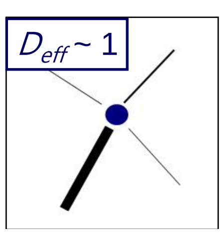
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)

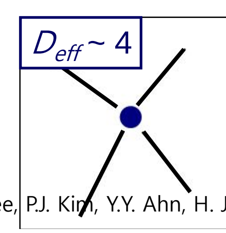
For $\alpha \to 1$, D_α is the "effective degree" D_{eff}

$$D_{eff}^{(i)} = \prod_j \tilde{w}_{ij}^{-\tilde{w}_{ij}} = \exp(-\sum_j \tilde{w}_{ij} \ln \tilde{w}_{ij}) = \exp(S) \le N$$

where S is the Shannon entropy

If weights are distributed **uniformly**, D_{eff} ~ degree k
If weights are highly **heterogeneous**, D_{eff} ~ 1

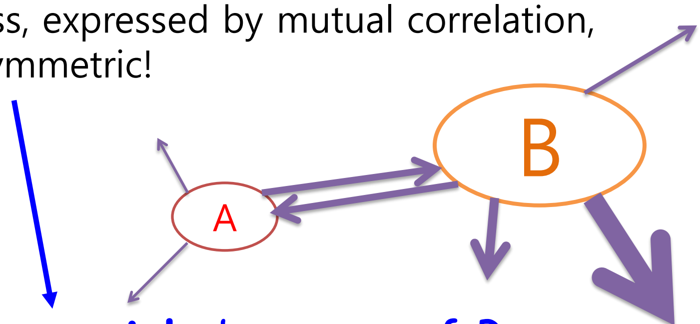




S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)

"hidden asymmetry" of relatedness

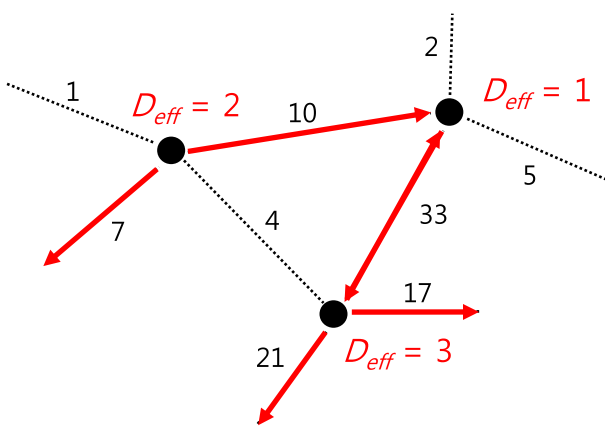
- Most social networks are undirected. (believed to be "mutual" relationship)
- Relatedness, expressed by mutual correlation, can be asymmetric!



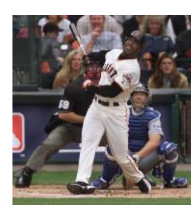
A is famous *mainly* because of B, but B is famous *not only* because of A!

S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)

weighted undirected network
→ directed network by D_{eff}



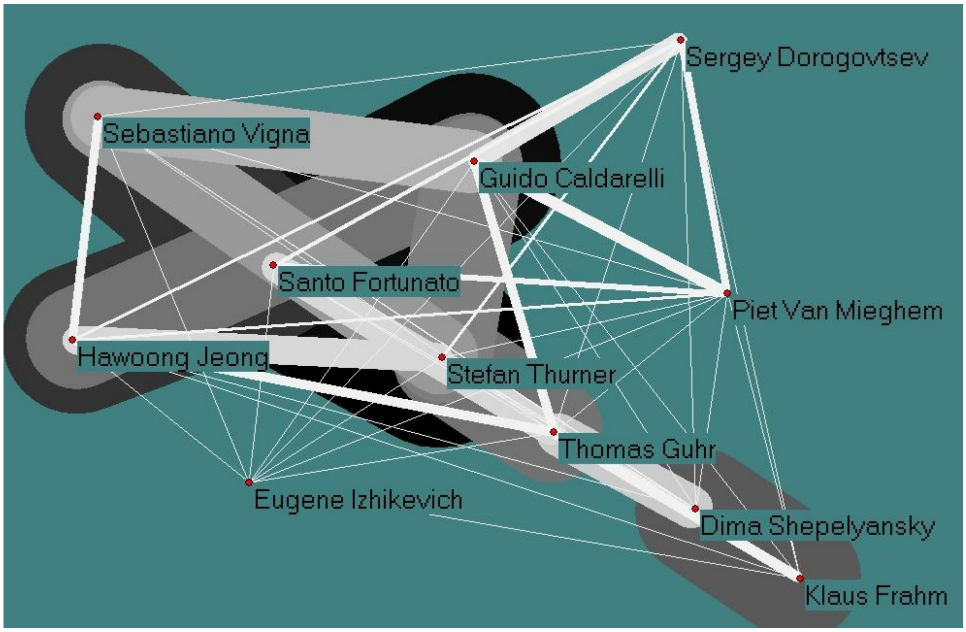
Maximum D_{in} (US Senate)



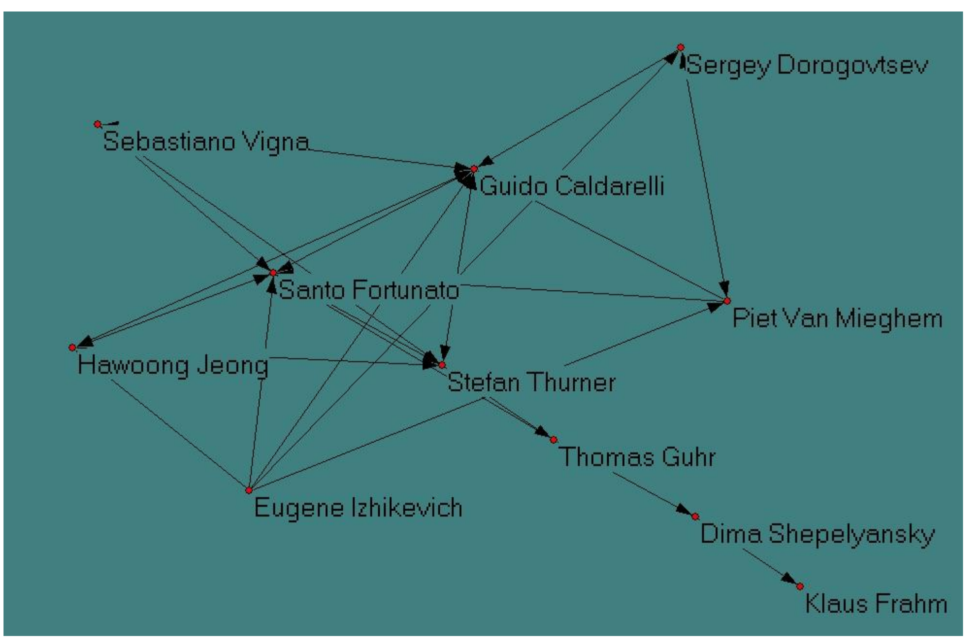
Maximum D_{in} (MLB players)

The more incoming links (D_{in}) a node has, the more influential the node is in the system!

S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (PLoS ONE 2010)



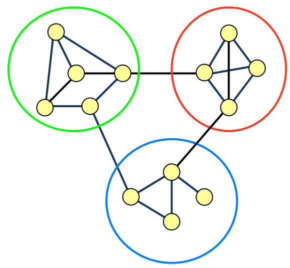
S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)



S.H. Lee, P.J. Kim, Y.Y. Ahn, H. Jeong (arXiv:0710.3268, PLoS ONE 2010)

More Specific Problems…

Community identification problem in directed networks:



General Definition:

More links are placed within communities and less links are placed between communities.

- Community structure is an ubiquitous property of many real-world networks, such as protein-protein interaction network, citation network, social relationship network, etc.

- Community structure is related to the structural and dynamical properties of many networks.

Y.D. Kim, S.W. Son, H. Jeong (PRE 2009)

Modularity

- Modularity Optimizing Methods

Find the maximum of modularity Q over possible community partitions of the network, and the partition of maximum modularity is taken as the best estimate of the communities in the network.

$Q =$ (fraction of links within communities) − (expected value of that fraction).

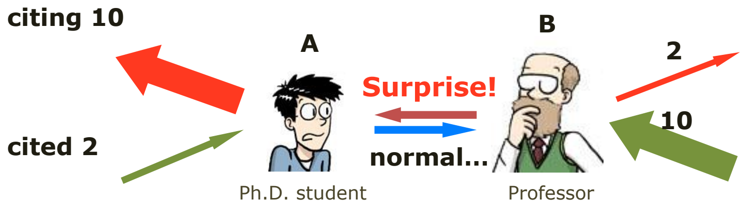
$$= \frac{1}{2M} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2M} \right] \delta_{c_i c_i}$$

where A_{ij} is the adjacency matrix, M is the total number of links, k_i is the degree of node *i*, and c_i is the community label of node *i*.

- However, modularity is only defined in undirected networks.

Newman and Girvan(2004) *Phys Rev E* , **69**, 026113

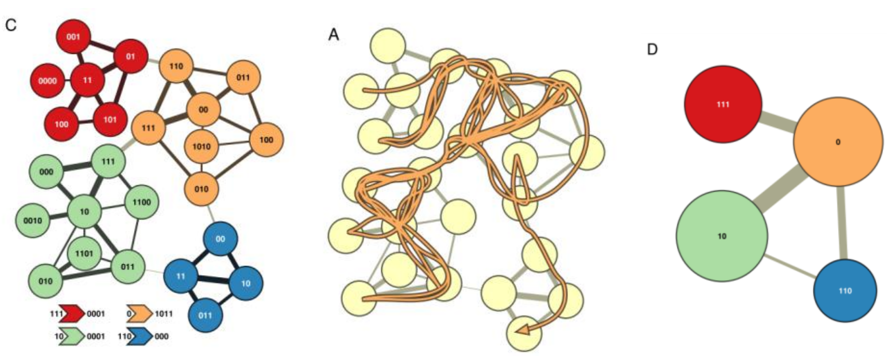
How to Consider Direction Information?



- a link from B to A implies stronger relation between A and B than a link from A to B.
- A pairs of nodes should be more likely to be included in the same community when the link is directing from more important node to less important node.

➔ Remind me of "PageRank"!

Let Random Walkers walk on directed network, and watch how long they stay in each nodes…

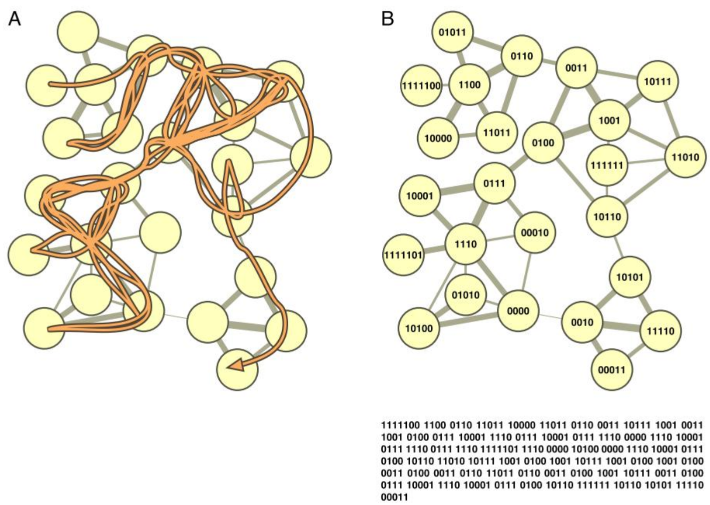


Picture from Rosvall2007

Rosvall2007

an information theoretic approach

Basic concept is the same. Detecting communities by compressing the description of information flows (random walk) on networks.



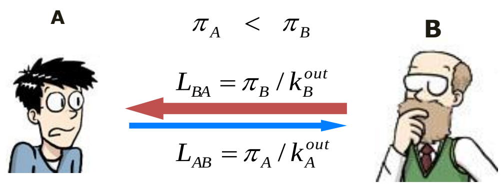
New Definitions:

Community — A group of nodes within which a random walker is more likely to stay.

Modularity — Q^{lr} = (fraction of **time** spent moving within communities) - (expected value of that fraction).

- $L_{ij} \equiv \pi_i G_{ij}$ **(LinkRank)**, the **probability** that a random walker found on the **link** ***i*** **→** ***j***.
- π_i **(PageRank)**, the **probability** that the random walker visiting **node** ***i***.
- $G_{ij} \equiv A_{ij} / k_i^{out}$ **(Google matrix)**, the **probability** that the random walker moves to **node** ***j*** when it is on node *i*. (Random hopping needs to be added if the network is not a Strongly Connected Component.)
- $\pi^T G = \pi^T$

- The **direction** effect is **properly** considered!



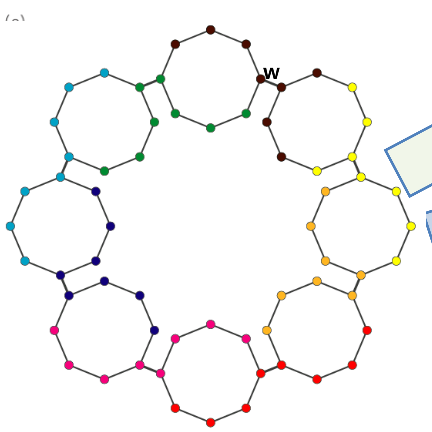
In undirected networks,

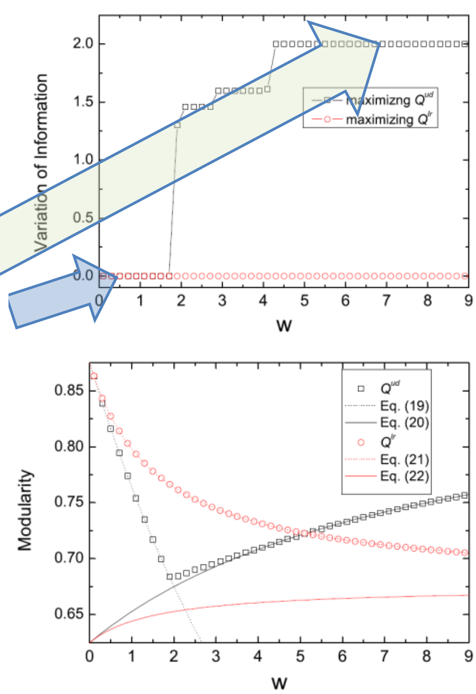
$$\pi_i = \frac{k_i}{2M}, \quad G_{ij} = \frac{A_{ij}}{k_i}. \quad \Rightarrow$$

$$Q^{lr} = \sum_{i,j} L_{ij}\delta_{c_i c_i} - \sum_{i,j} E(L_{ij})\delta_{c_i c_j} = \sum_{i,j}\left[\frac{A_{ij}}{2M} - \frac{k_i}{2M}\frac{k_j}{2M}\right]\delta_{c_i c_i} = Q^{ud}.$$

- The new definitions **consist well** with the original ones.

Results (m=n=8)





Dynamics on Directed Network

Synchronization

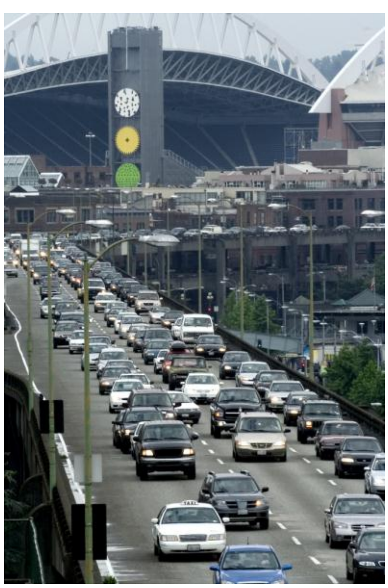
- Cristiaan Huygens, 1656
- Syn (the same) + chro (time) + nize : agreement in time.
- One of the most fascinating collective behaviors in nature.
- A unit interacts with other units, and tries to imitate other's behavior.
- Fireflies, cricket, neurons, and cardiac pacemaker cells.
- Coupled oscillators. Kuramoto model (1975)

$$\frac{d\phi_i}{dt} = \omega_i - \frac{K}{\langle k \rangle}\sum_{j=1}^{N} a_{ij}\sin(\phi_i - \phi_j)$$


S.W. Son, B.J. Kim, H.Hong, H. Jeong PRL (2009)

Motivation

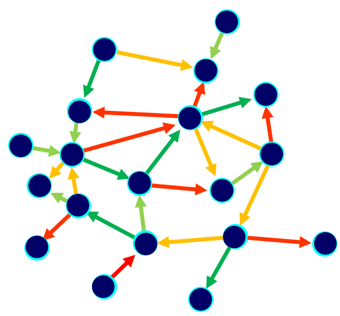
Traffic



Parallel computing

Internet
Power grid

Effect of link directions

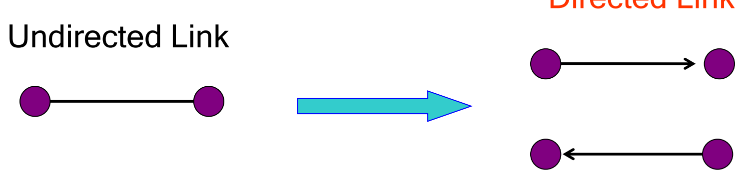


Randomly ?

Q. How can we enhance (or control) the synchronization on complex network?

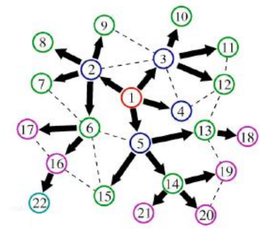
Directionality

Visiting every link…



Maximum Synchronizability

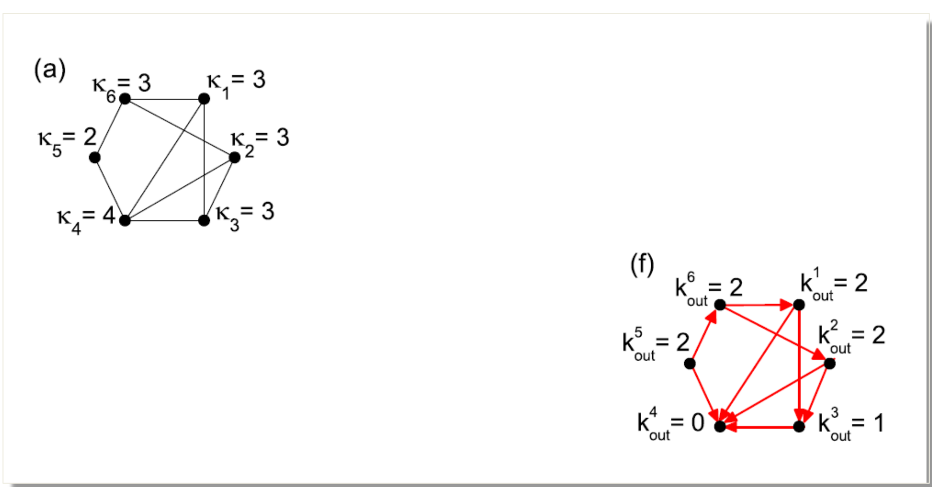
Nishikawa analytically shows that…



“ We first note that **maximum synchronizability** can always be achieved by imposing that the network (i) embeds an oriented spanning tree, (ii) has no directed loops, and (iii) has normalized input strengths in each node, i.e., the total input is the same for all nodes that have input. ”

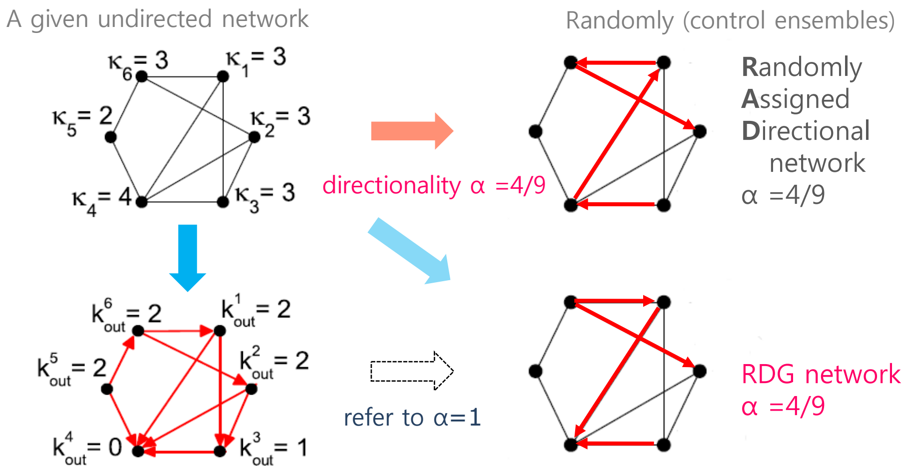
T. Nishikawa and A. E. Motter, Phys. Rev. E **73**, 065106(R) (2006).

Residual Degree Gradient (RDG) Network



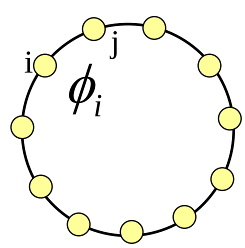
κ_i : residual degree - Number of remaining links without assigning direction.

RDG vs. RAD with directionality α



RDG with α =1. (Reference)

Kuramoto model



N coupled limit-cycle oscillators

$$\frac{d\phi_i}{dt} = \omega_i - \frac{K}{\langle k \rangle} \sum_{j=1}^{N} a_{ij} \sin(\phi_i - \phi_j)$$

phase

$$\{ \phi_i(t) \mid i = 1, 2, \ldots, N \}$$

1.Natural frequency

$$g(\omega) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\omega^2}{2}}$$

2.Adjacency matrix

$$a_{ij} \neq a_{ji}$$

3.Coupling strength

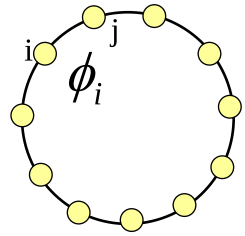
K > 0

4.Mean degree

$$\langle k \rangle = \frac{\sum_{i=1}^{N} k_i}{N} = \frac{\sum_{i,j=1}^{N} a_{ij}}{N}$$

Y. Kuramoto, in *International Symposium on Mathematical Problems in Theoretical Physics*, edited by H. Araki, Lecture Notes in Physics Vol. 39 (Springer-Verlag, Berlin, 1975)

Kuramoto model



N coupled limit-cycle oscillators

$$\frac{d\phi_i}{dt} = \omega_i - \frac{K}{\langle k \rangle}\sum_{j=1}^{N} a_{ij}\sin(\phi_i - \phi_j)$$

phase

$$\{\phi_i(t) \mid i = 1, 2, \ldots, \mathrm{N}\}$$

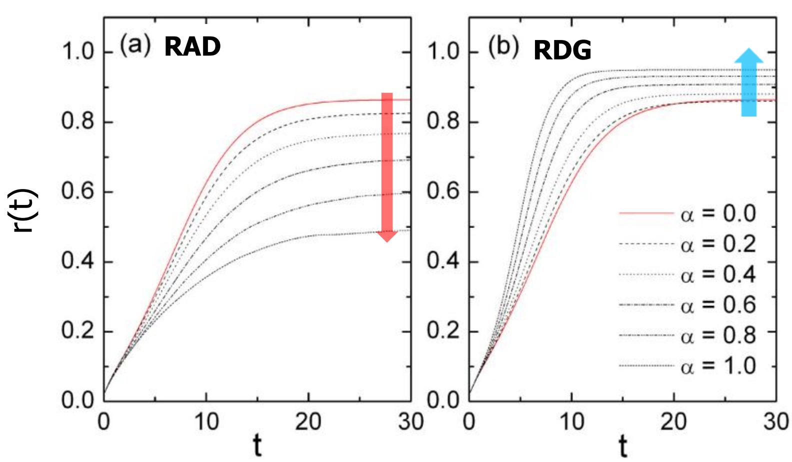
Order parameter

$$r(t) \equiv \left\langle \left| \frac{1}{N}\sum_{j=1}^{N} e^{i\phi_j(t)} \right| \right\rangle_{ensemble}$$

Y. Kuramoto, in *International Symposium on Mathematical Problems in Theoretical Physics*, edited by H. Araki, Lecture Notes in Physics Vol. 39 (Springer-Verlag, Berlin, 1975)

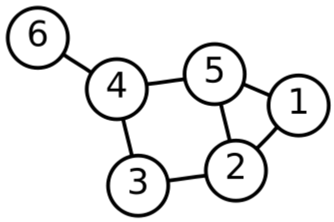
Kuramoto model

SW network, N=1600, P=0.2, <k>=6



α: directionality

Eigenvalues of Laplacian Matrix



Adjacency matrix

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

Laplacian matrix (positive semidefinite)

$$G = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$$

$$G_{ij} = -A_{ij} + \delta_{ij} \sum_{j=1}^{N} A_{ij}$$

Eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$$

Smallest nonzero eigenvalue

Largest eigenvalue

Laplacian matrix is related with diffusion process and network community.

$$\dot{\mathbf{x}}_i = \mathrm{F}(\mathbf{x}_i) + \sigma \sum_{j=1}^{N} A_{ij} \left[\mathrm{H}(\mathbf{x}_j) - \mathrm{H}(\mathbf{x}_i) \right]$$

$$= \mathrm{F}(\mathbf{x}_i) - \sigma \sum_{j=1}^{N} G_{ij}\, \mathrm{H}(\mathbf{x}_j)$$

$$G_{ij} = -A_{ij} + \delta_{ij} \sum_{j=1}^{N} A_{ij} = L_{ij}$$

$$0 = \lambda_1 \leq \mathrm{Re}\,\lambda_2 \leq \cdots \leq \mathrm{Re}\,\lambda_N$$

Eigenratio

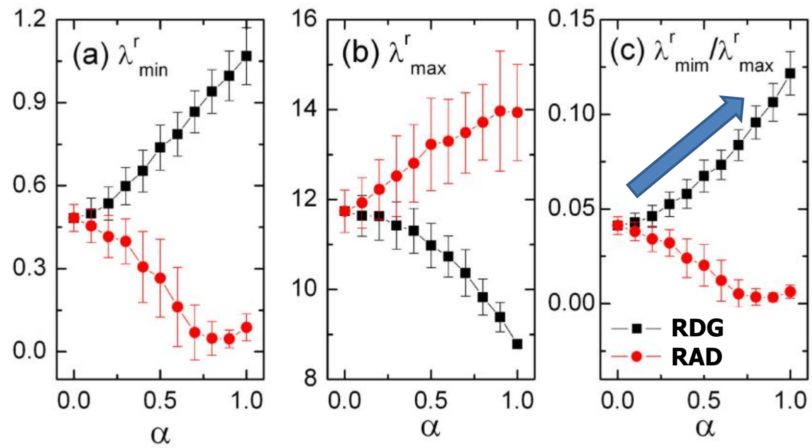
$$R = \lambda^r_{\mathrm{min}} / \lambda^r_{\mathrm{max}}$$

"The larger the eigenratio R,
the larger the synchronizability of the network
and vice versa."

T.Nishikawa et. al., Phys. Rev. Lett. **91**, 014101 (2003).

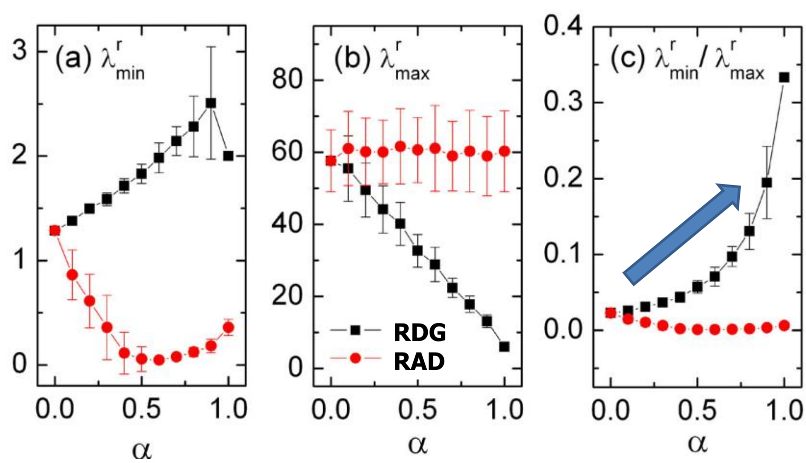
Eigenvalue Ratio

SW network, N=400, P=0.2, <k>=6

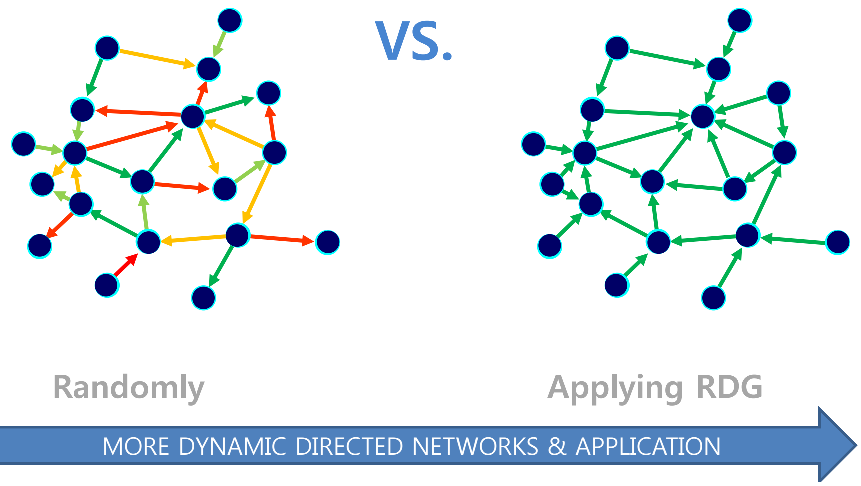


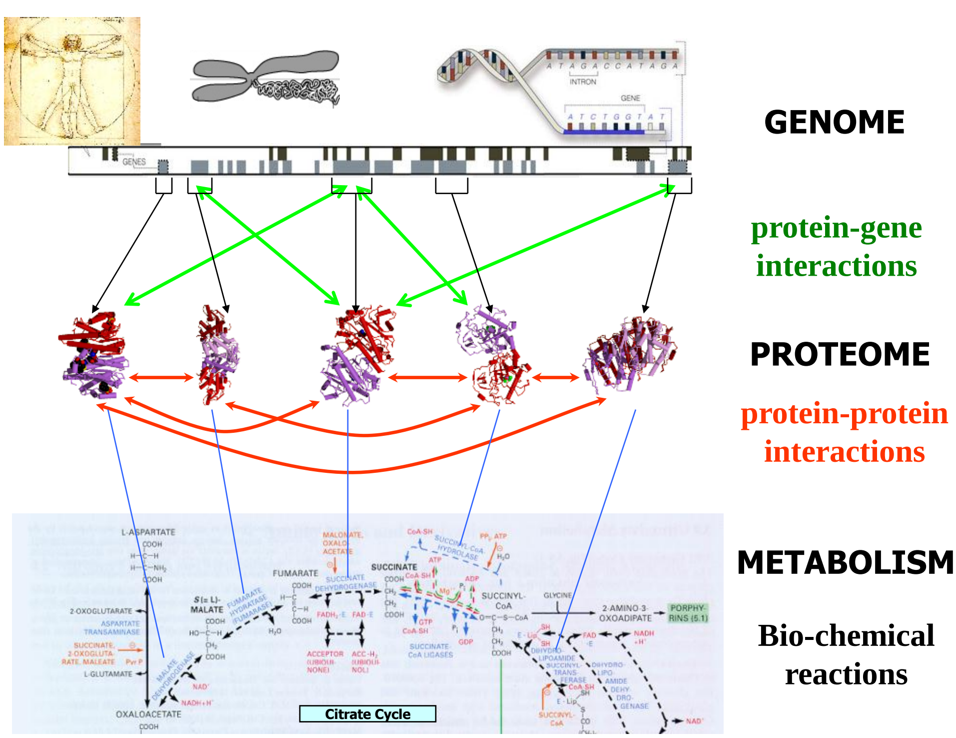
Eigenvalue Ratio

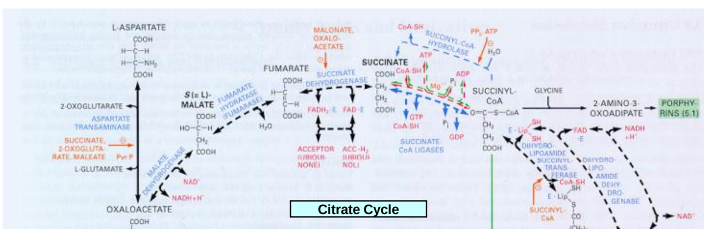
SF network, N=400, <k>=6



Smart link directions can enhance synchronization!

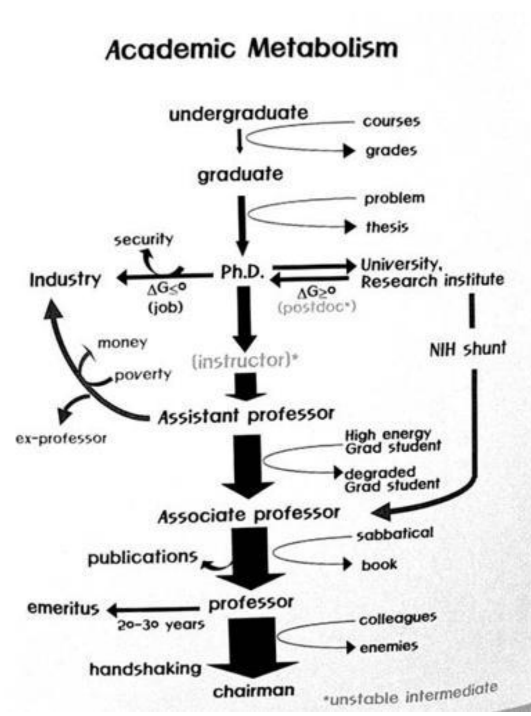




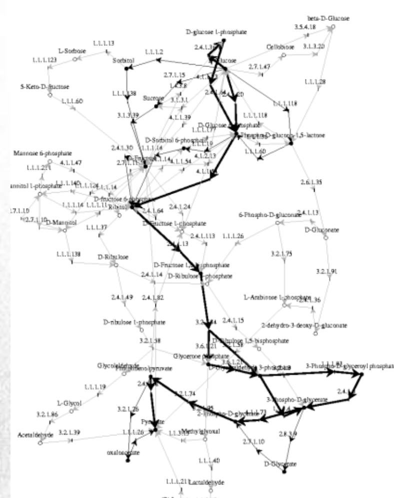


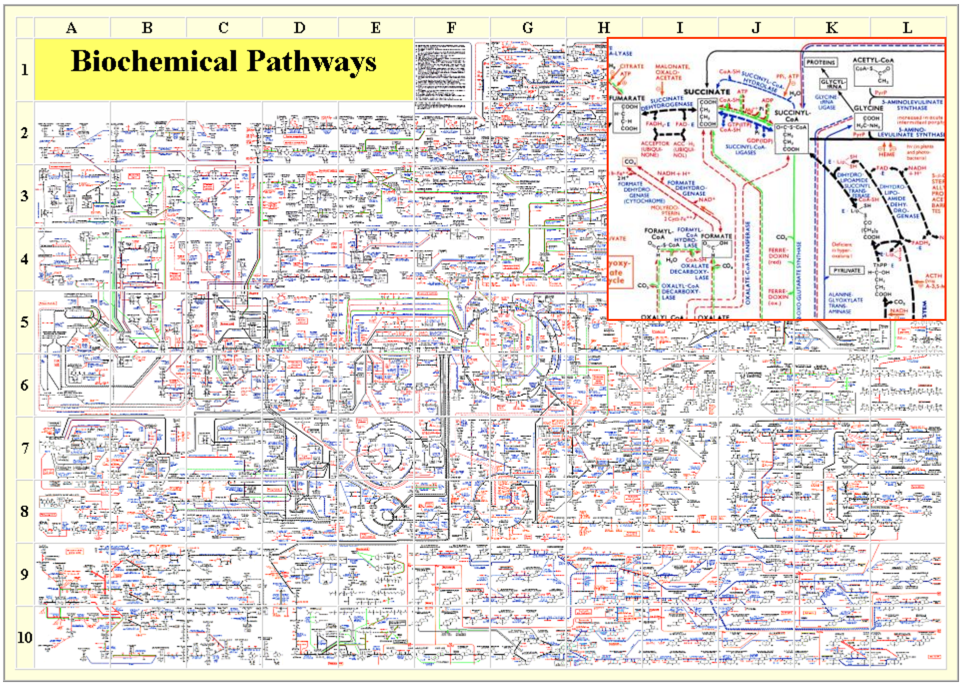
Metabolic Networks

Bio-chemical reactions



- **Nodes**: chemicals (proteins, substrates)
- **Links**: bio-chem. reaction



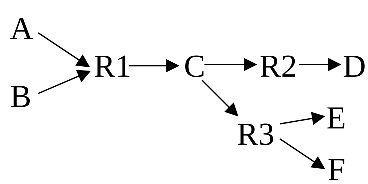


Construction of metabolic pathway as a graph

$$A + B \xrightarrow{R1} C \xrightarrow{R2} D \qquad C \xrightarrow{R3} E + F$$

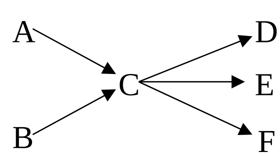
1. Bipartite network

- Two kinds of nodes: substrates and reactions



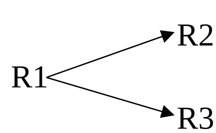
2. Substrate to substrate network

- Node: substrate. Link: chemical reaction

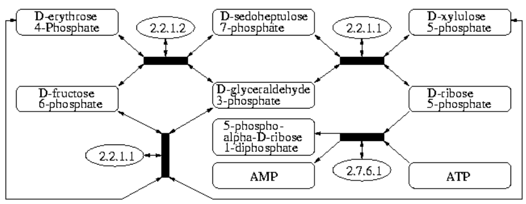


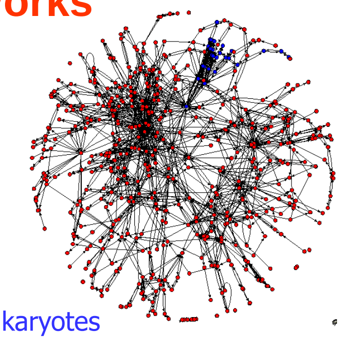
3. Reaction to reaction network

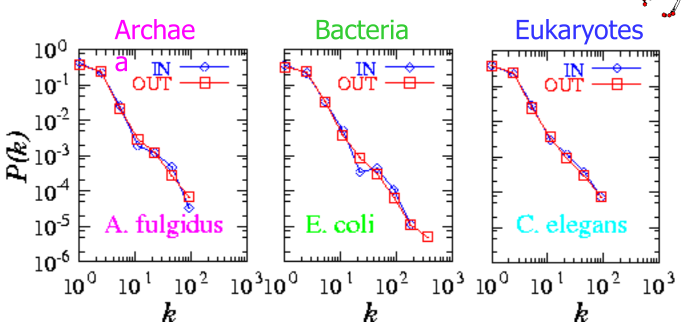
- Node: reaction. Link: substrate



Metabolic networks





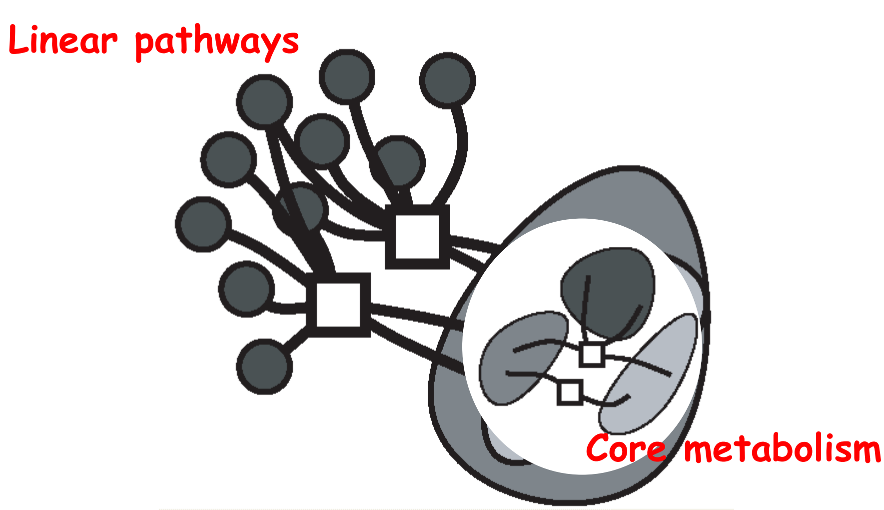


Organisms from all three domains of life are **scale-free** networks!

$P(k) \sim k^{-\gamma}$

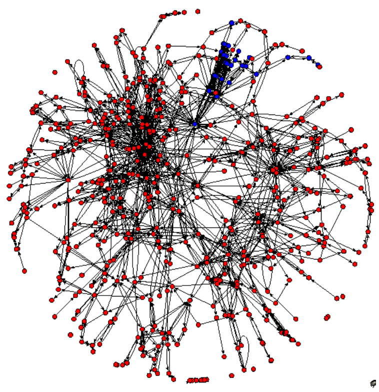
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature, **407** 651 (2000)

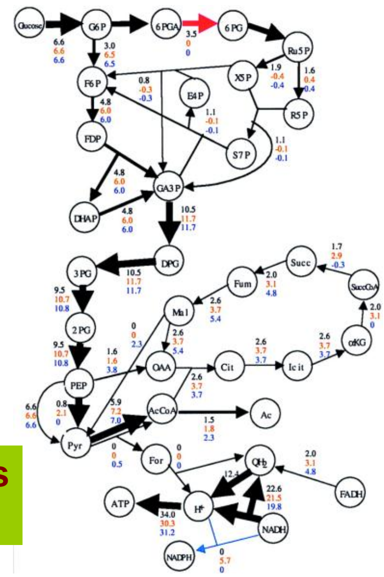
Schematic picture of metabolic networks



P. Holme, M. Huss, H. Jeong, Bioinformatics (2003)

All these works were about STATIC topological properties of metabolic networks

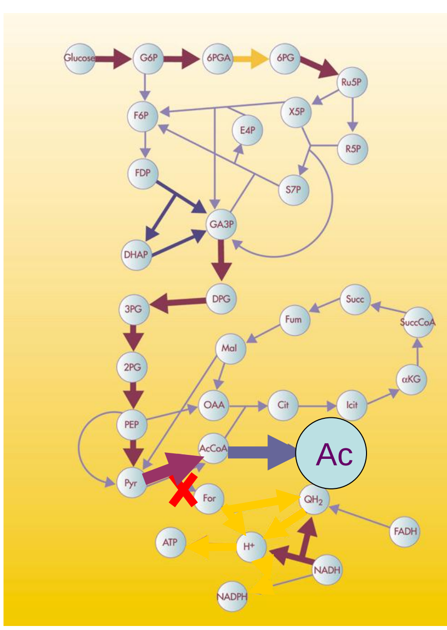




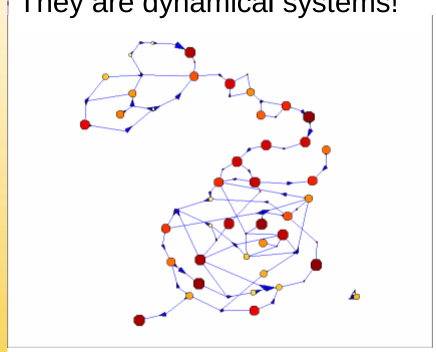
But real metabolic networks are DYNAMIC!!

Metabolic Engineering

Complex network **research center**



Metabolic networks are not static!
They are dynamical systems!



Q: How to control the metabolic networks?
e.g.
How to increase product Ac?

To fully understand the dynamics of metabolism inside a cell

Requirements

- Complete knowledge of all the biochemical pathways in a cell, i.e. complete knowledge of all the possible stoichiometric reactions in a cell
- Enzyme kinetics
- Regulation of all enzymes in a cell

Problems

- Complete kinetic and regulatory control parameters are not available for all the enzymes in a cell.
- The kinetic constants are based on *in vitro* conditions which are almost completely different from the *in vivo* conditions

To fully understand the dynamics of metabolism inside a cell

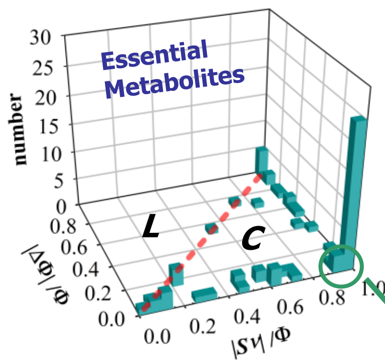
Approaches

- Complete knowledge of all the biochemical pathways in a cell, i.e. complete knowledge of all the possible stoichiometric reactions in a cell
- ~~Enzyme kinetics~~
- ~~Regulation of all enzymes in a cell~~

Flux Balance Analysis

- Linear metabolic model based on stoichiometric balance equations in a cell. (mass conservation)
- Steady state solution. (no net accumulation of metabolites)
- Uses linear optimization techniques to study the flux distribution in a cell. (with appropriate objective function)

Eliminating the maximum $\frac{|S_{ij}v_j|}{FS_i}$



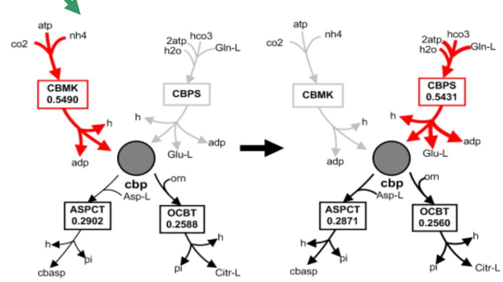
Most essential metabolites are situated in region C (compensation) rather than L (loss).

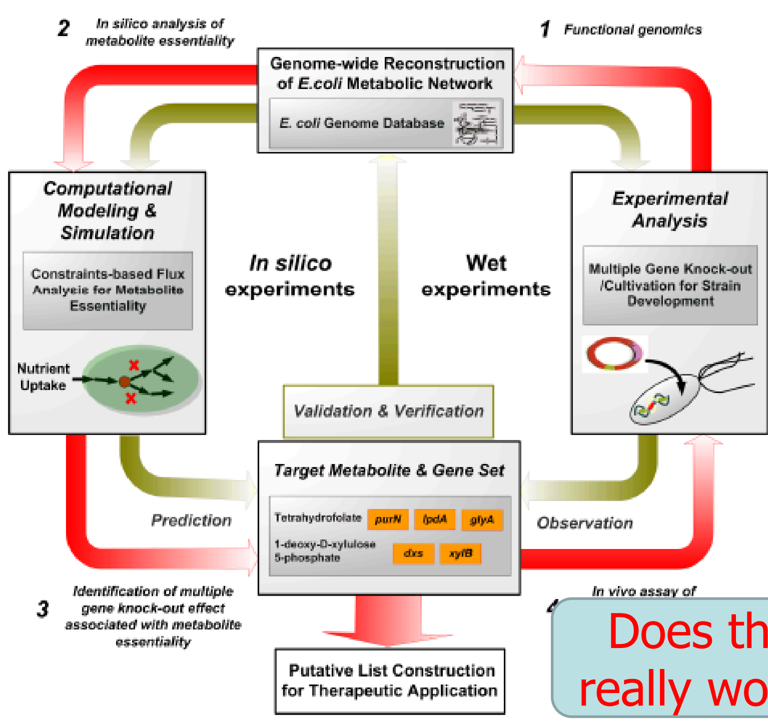
$$C:\ \left.\frac{|S_{ij}v_j|}{FS_i}\right|_{\max} > \frac{|\Delta FS_i|}{FS_i}$$

If not, L.

Carbamoyl phosphate (cbp):

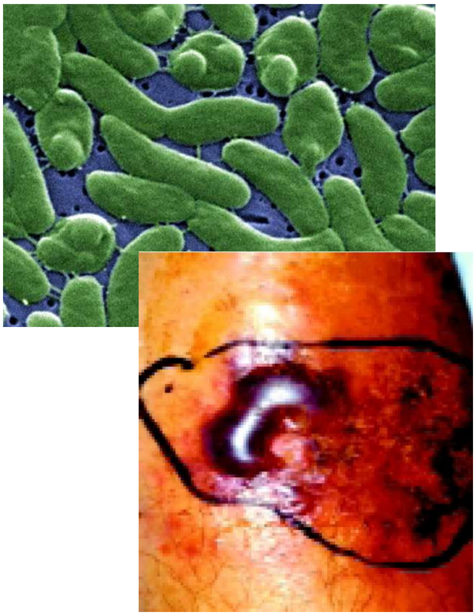
98.9% of the flux-sum is recovered when the highest flux of carbamate kinase is eliminated.



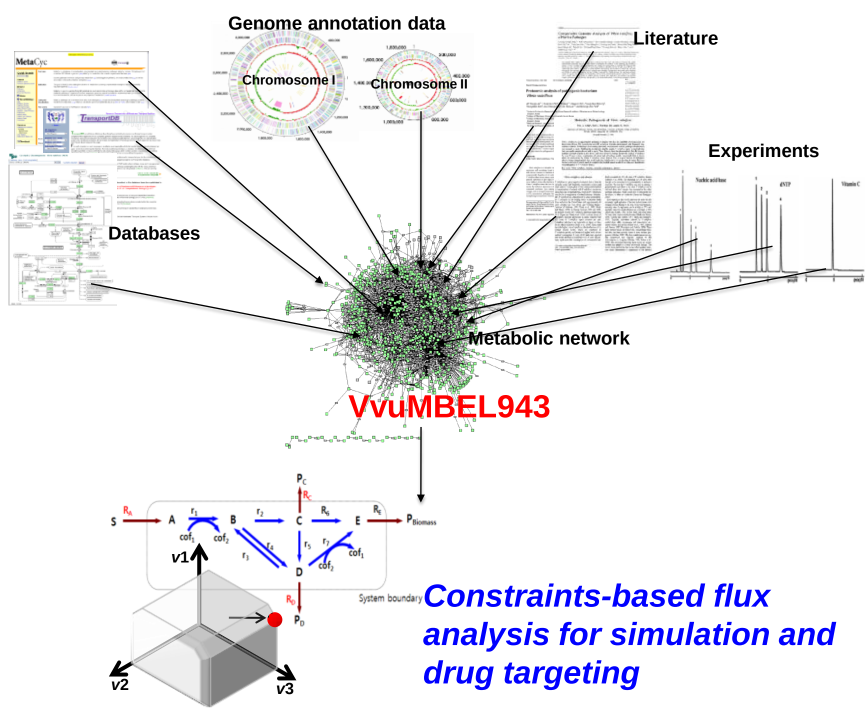


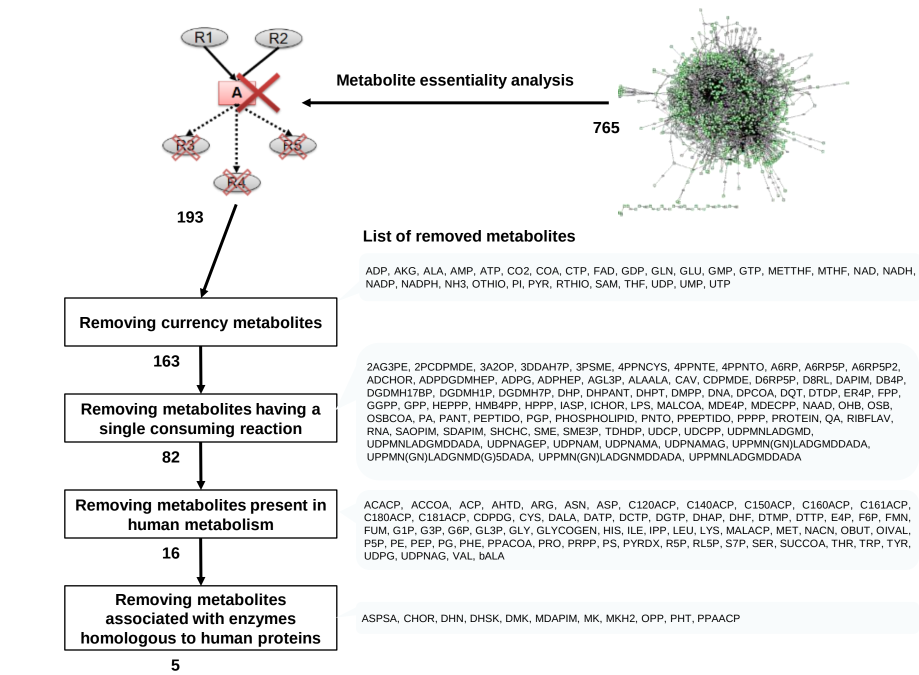
Does this really work?

Characteristics of *Vibrio vulnificus* CMCP6

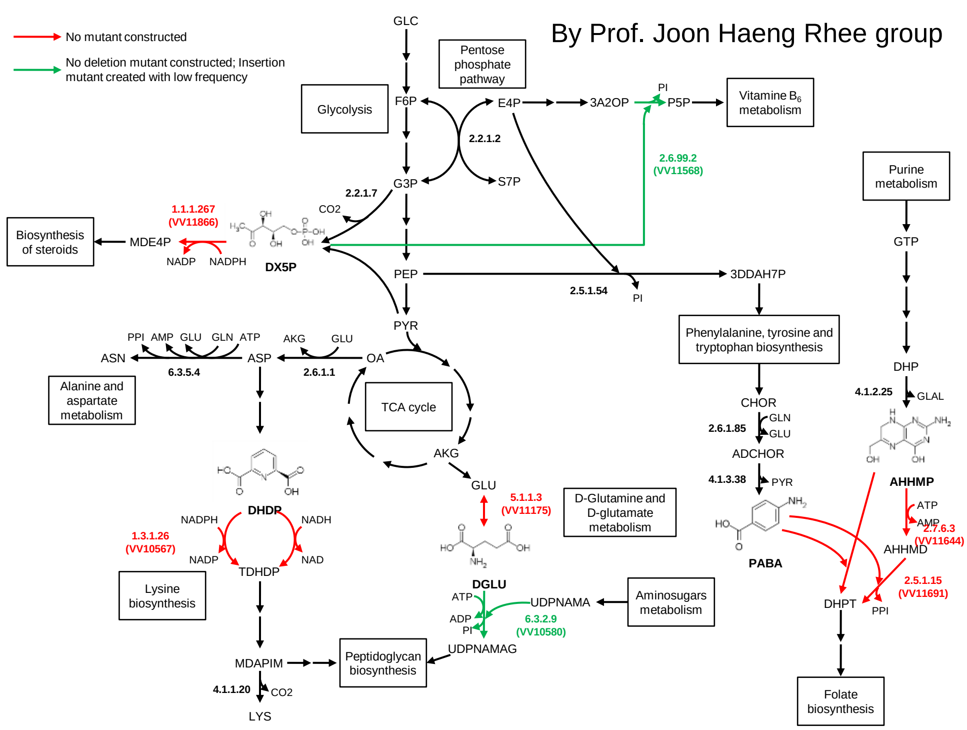


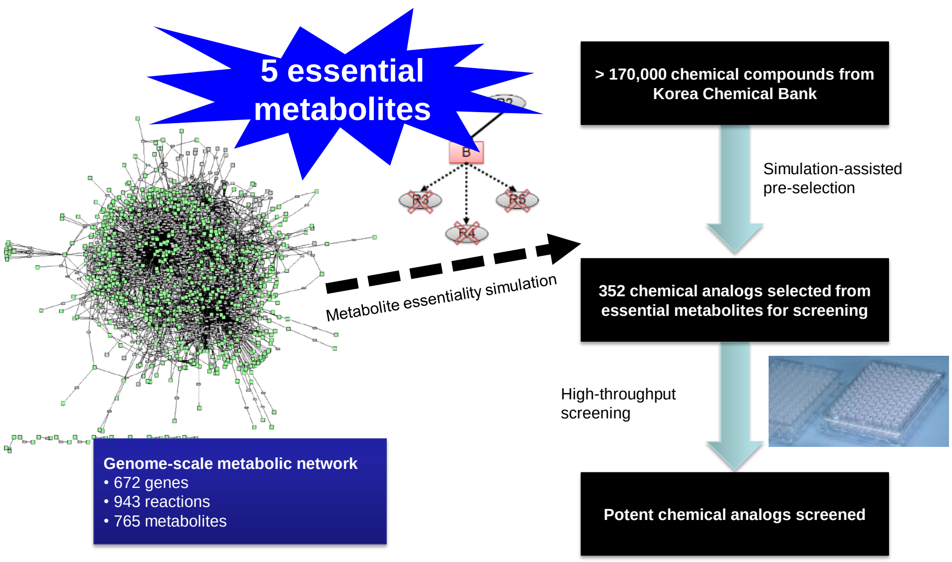
- ✓ Gram-negative bacterium
- ✓ Biotype 1: predominant human pathogens
- ✓ Typically found in estuarine waters
- ✓ Currently two strains' genomes sequenced: YJ016 (Chen *et al. Genome Res.* **13**, 2577-2587, 2003) and CMCP6 (not reported)
- ✓ Genome size: 5.12 Mb
- ✓ Causing agent of septicemia, necrotizing wound infection, and gastroenteritis

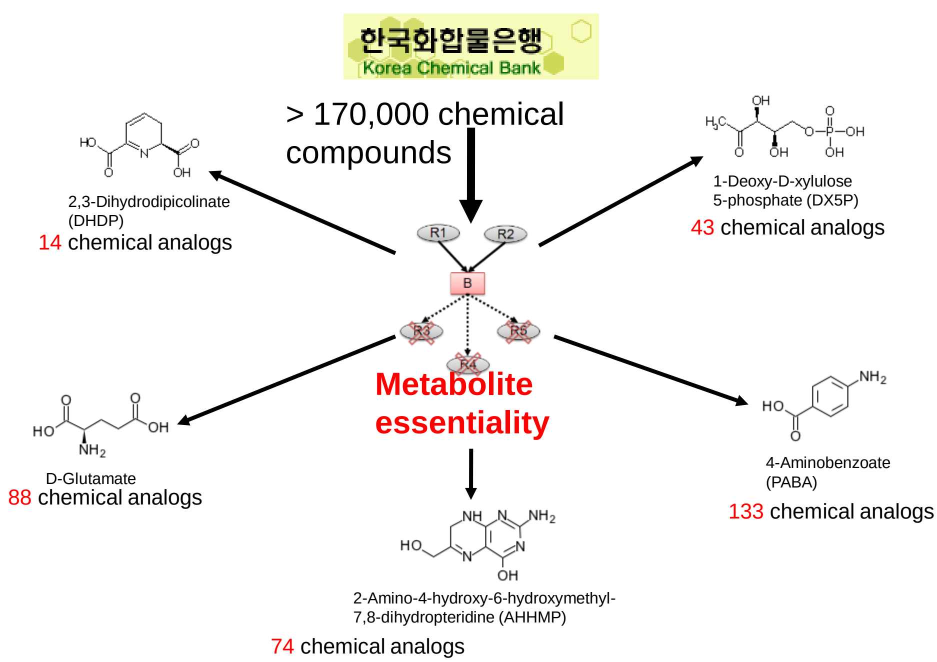




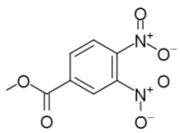
By Prof. Joon Haeng Rhee group



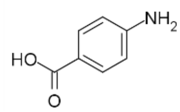

5 essential metabolites
B
R3
R4
R5
Metabolite essentiality simulation
Genome-scale metabolic network
• 672 genes
• 943 reactions
• 765 metabolites
> 170,000 chemical compounds from Korea Chemical Bank
Simulation-assisted pre-selection
352 chemical analogs selected from essential metabolites for screening
High-throughput screening
Potent chemical analogs screened


한국화합물은행
Korea Chemical Bank
> 170,000 chemical compounds
R1
R2
B
R3
R4
R5
Metabolite essentiality
2,3-Dihydrodipicolinate (DHDP)
14 chemical analogs
1-Deoxy-D-xylulose 5-phosphate (DX5P)
43 chemical analogs
D-Glutamate
88 chemical analogs
4-Aminobenzoate (PABA)
133 chemical analogs
2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine (AHHMP)
74 chemical analogs

Final 'hit', the most potent analog



24837
(Chemical analog)

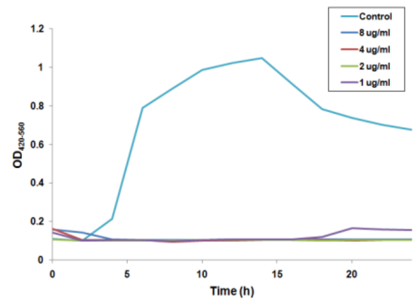


4-Aminobenzoate
(PABA, metabolite)

Compound ID	MIC (μg/ml)	MBC (μg/ml)
24837	**2**	**4**

•MIC: minimal inhibitory concentration
•MBC: minimal bactericidal concentration

Growth Inhibition



Final essential metabolites of *A. baumannii* AYE

Essential metabolites	ORFs	Metabolism	E.C. Number	Enzyme
AHHMP	ABAYE1418 OR ABAYE3176	Folate biosynthesis	2.7.6.3	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase
	ABAYE0807 OR ABAYE3568 OR ABAYE3612 OR ABAYE3616	Folate biosynthesis	2.5.1.15	dihydropteroate synthase
DGLU	ABAYE0082 OR ABAYE3395	D-Glutamine and D-glutamate metabolism	5.1.1.3	glutamate racemase
	ABAYE3524	Peptidoglycan biosynthesis	6.3.2.9	UDP-N-acetylmuramoylalanine--D-glutamate ligase
DHDP	ABAYE0036	Lysine biosynthesis	1.3.1.26	dihydrodipicolinate reductase
DHP	ABAYE1417	Folate biosynthesis	4.1.2.25	dihydroneopterin aldolase
	ABAYE0811	Folate biosynthesis	3.1.3.1	alkaline phosphatase D precursor
DHSK	ABAYE1539 OR ABAYE1682	Phenylalanine, tyrosine and tryptophan biosynthesis	4.2.1.10	3-dehydroquinate dehydratase II; catabolic 3-dehydroquinate dehydratase (3-dehydroquinase)
	ABAYE0377	Phenylalanine, tyrosine and tryptophan biosynthesis	1.1.1.25	shikimate 5-dehydrogenase
	ABAYE1685	Phenylalanine, tyrosine and tryptophan biosynthesis	1.1.99.25	quinate/shikimate dehydrogenase
	ABAYE1683	Phenylalanine, tyrosine and tryptophan biosynthesis	4.2.1.-	3-dehydroshikimate dehydratase
DX5P	ABAYE1581	Biosynthesis of steroids	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
	ABAYE0945	Vitamin B6 metabolism	2.6.99.2	pyridoxine 5-phosphate synthase
DQT	ABAYE1539 OR ABAYE1682	Phenylalanine, tyrosine and tryptophan biosynthesis	4.2.1.10	3-dehydroquinate dehydratase II; catabolic 3-dehydroquinate dehydratase (3-dehydroquinase)
	ABAYE1685	Phenylalanine, tyrosine and tryptophan biosynthesis	1.1.99.25	quinate/shikimate dehydrogenase
KDO	ABAYE2076	Lipopolysaccharide biosynthesis	2.7.7.38	3-deoxy-manno-octulosonate cytidylyltransferase
	-	Lipopolysaccharide biosynthesis	-	Hypothetical reaction that consumes KDO to genera
PABA	ABAYE0807 OR ABAYE3568 OR ABAYE3612 OR ABAYE3616	Folate biosynthesis	2.5.1.15	dihydropteroate synthase

AHHMP, 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine; **DGLU**, D-Glutamate; **DHDP**, 2,3-Dihydrodipicolinate; **DHP**, 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8-dihydropteridine; **DHSK**, 3-Dehydroshikimate; **DX5P**, 1-Deoxy-D-xylulose 5-phosphate; **DQT**, 3-Dehydroquinate; **KDO**, 2-Dehydro-3-deoxy-D-octonate; **PABA**, 4-Aminobenzoate.

Kim et al. *Mol. BioSyst.*, 6, 339-348 (2010)

Summary

- There are many “directed” networks & interesting properties!
- Undirected (weighted) network can be mapped into directed network using “effective degree”.
  (arXiv:0710.3268, PLoS ONE 2010)
- LinkRank: one way of detecting communities on directed networks (PRE 2009)
- Increasing the directionality may enhance the network synchronization if follow the residual degree gradient method contrary to the result of *randomly assigned direction* (PRL 2009)
- Biological application to find drug target via network dynamics(FBA) (PNAS 2007)

Complex Systems & Statistical Physics Lab

http://stat.kaist.ac.kr