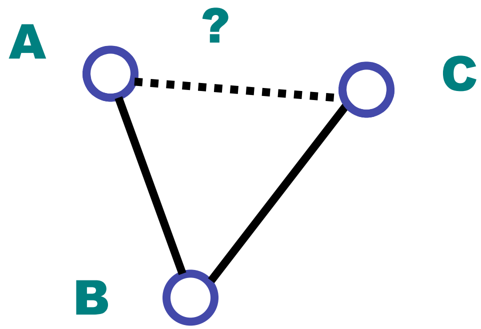
Transitivity, triadic closure, clustering

- Transitivity:
  - if A is connected to B and B is connected to C what is the probability that A is connected to C?
  - my friends' friends are likely to be my friends



Clustering

- Global clustering coefficient

  3 x number of triangles in the graph

  number of connected triples of vertices

$$C = \frac{3 \times \text{number of triangles in the graph}}{\text{number of connected triples}}$$

Local clustering coefficient (Watts&Strogatz 1998)

- For a vertex i
  - The fraction pairs of neighbors of the node that are themselves connected
  - Let n_i be the number of neighbors of vertex *i*

$$C_i = \frac{\text{\# of connections between i's neighbors}}{\text{max \# of possible connections between i's neighbors}}$$

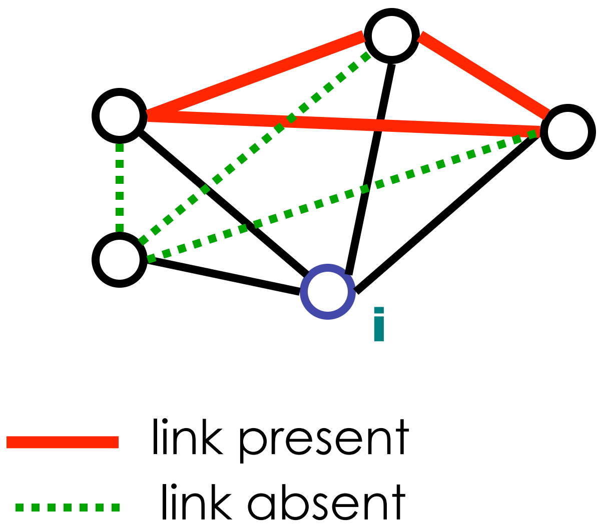
$$C_{i\ \text{directed}} = \frac{\text{\# directed connections between i's neighbors}}{n_i * (n_i - 1)}$$

$$C_{i\ \text{undirected}} = \frac{\text{\# undirected connections between i's neighbors}}{n_i * (n_i - 1)/2}$$

Local clustering coefficient (Watts&Strogatz 1998)

- Average over all n vertices

$$C = \frac{1}{n} \sum_i C_i$$



$n_i = 4$

max number of connections:

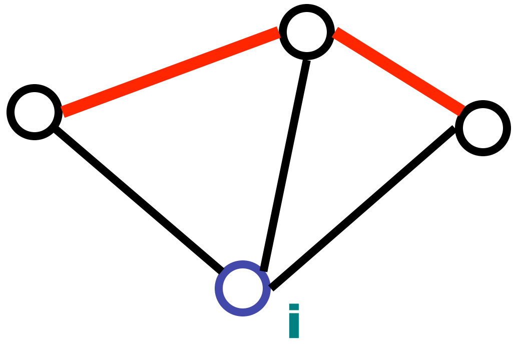
4*3/2 = 6

3 connections present

$C_i = 3/6 = 0.5$

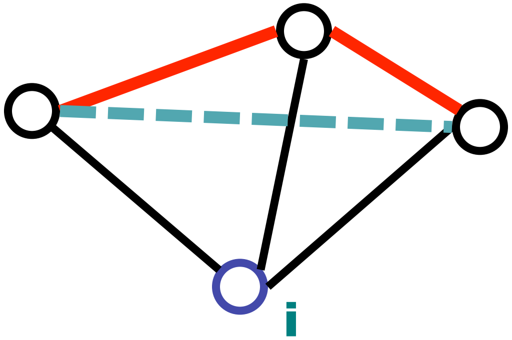
Quiz Q:

- The clustering coefficient for vertex A is:



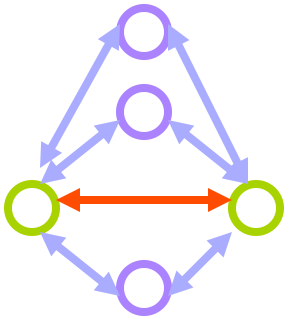
Explanation

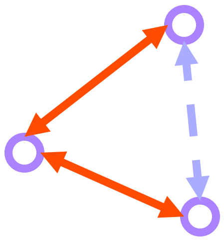
- $n_i = 3$
- there are 2 connections present out of max of 3 possible
- $C_i = 2/3$



Are strong ties "local"?

- A strong tie
  - frequent contact
  - affinity
  - many mutual contacts

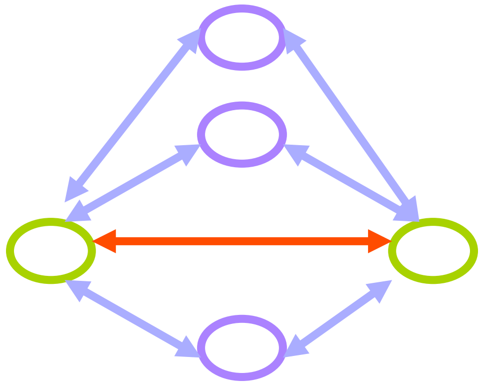




"forbidden triad":
strong ties are likely to "close"

edge embeddeness

- embeddeness: number of common neighbors the two endpoints have

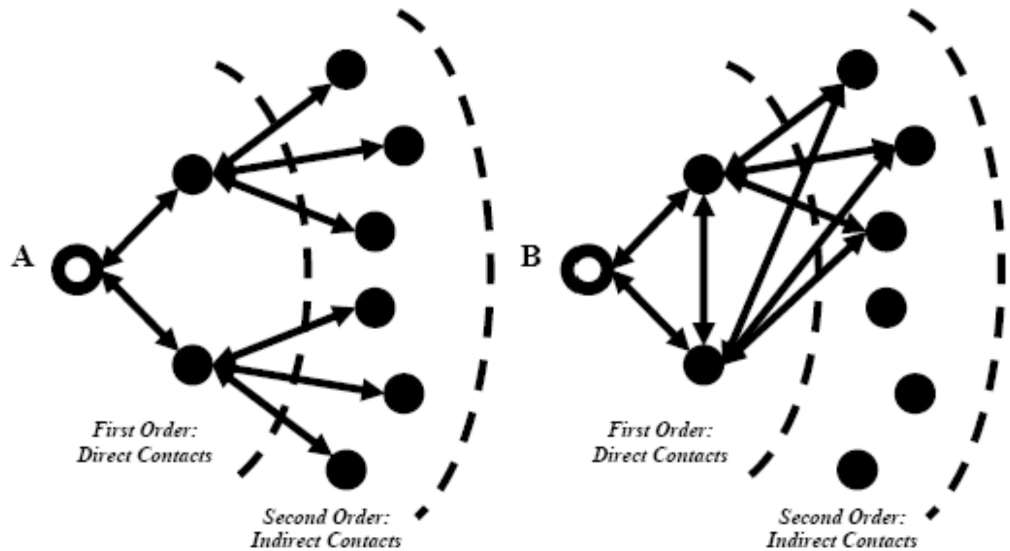


- neighborhood overlap:

$$\frac{\text{number of nodes who are neighbors of } \textit{both } A \text{ and } B}{\text{number of nodes who are neighbors of } \textit{at least one of } A \text{ or } B}$$

school kids and 1st through 8th choices of friends

- snowball sampling:
  - will you reach more different kids by asking each kid to name their 2 best friends, or their 7th & 8th closest friend?



Source: M. van Alstyne, S. Aral. Networks, Information & Social Capital, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=958158

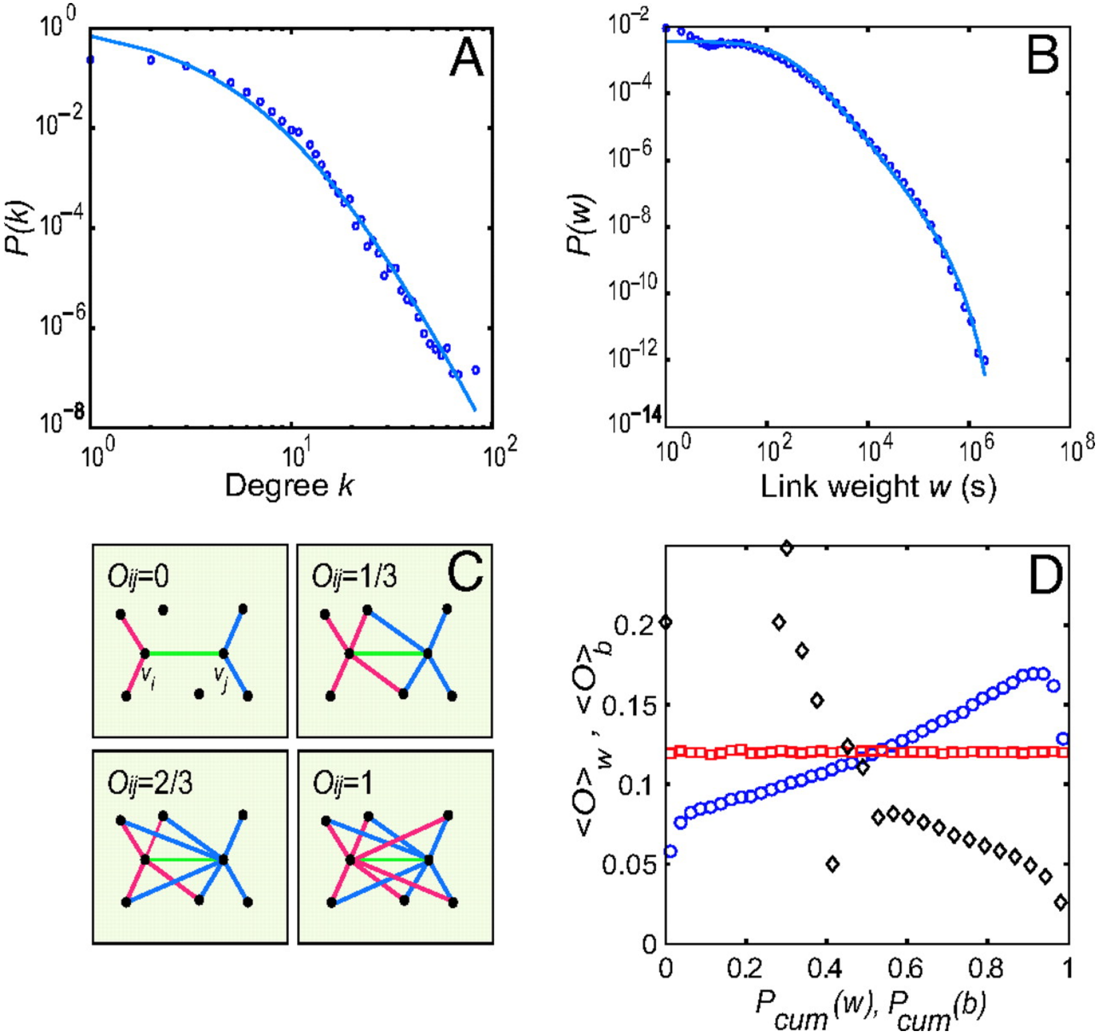
is it good to be embedded?

- What are the advantages of occupying an embedded position in the network?
- What are the disadvantages of being embedded?
- Advantages of being a broker (spanning structural holes)?
- Disadvantages of being a broker?

the strength of intermediate ties

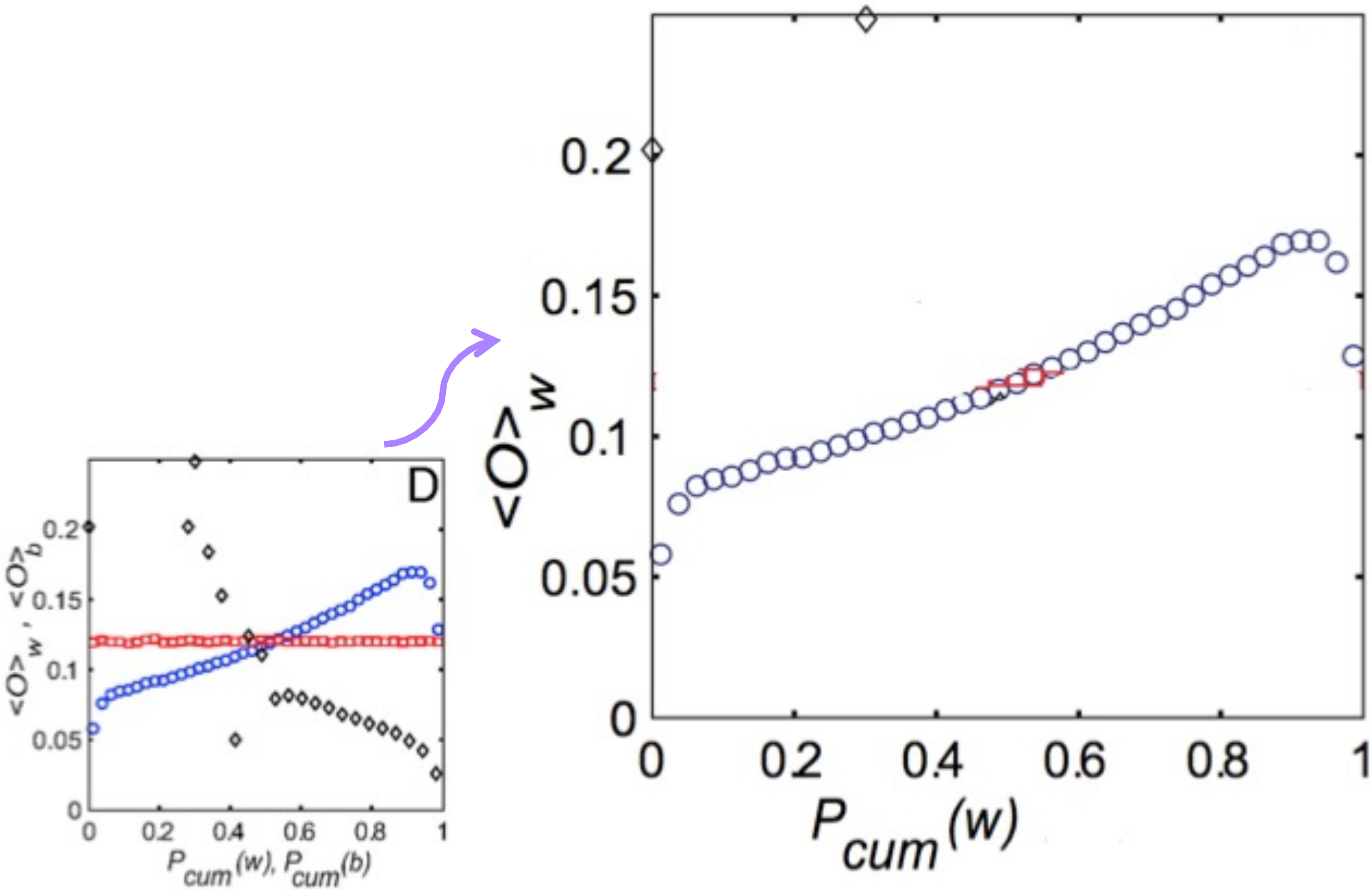
- study of a large call graph
- strong ties
  - frequent communication, but ties are redundant due to high clustering
- weak ties
  - reach far across network, but communication is infrequent…
- Onnela J. et.al. PNAS 2007;104:7332-7336
  - use nation-wide cellphone call records and simulate diffusion using actual call timing
    - in simulation, individuals are most likely to obtain novel information through ties of intermediate strength

Characterizing the large-scale structure and the tie strengths of the mobile call graph

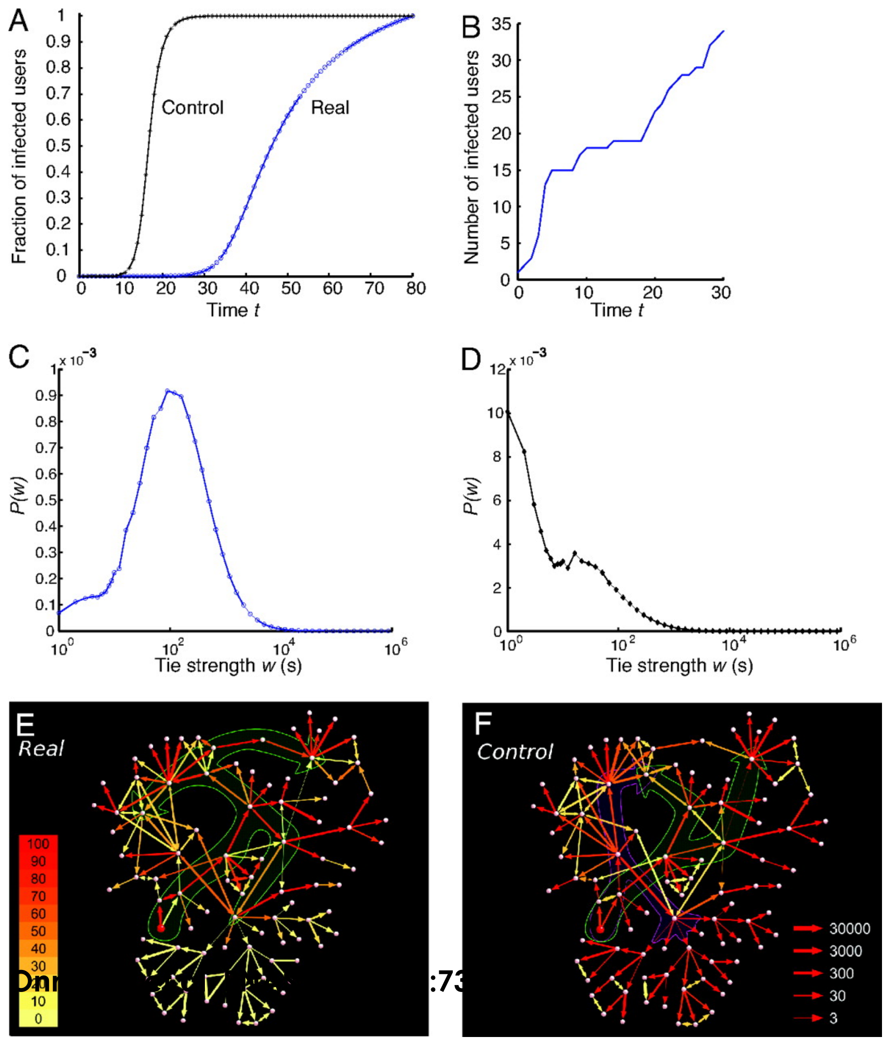


Onnela J et al. PNAS 2007;104:7332-7336

Edge neighborhood overlap as a function of tie strength

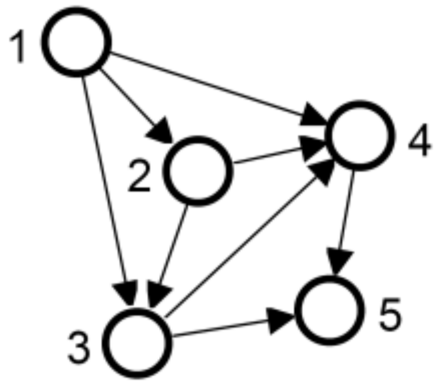


The dynamics of spreading on the weighted mobile call graph, assuming that the probability for a node v_i to pass on the information to its neighbor v_j in one time step is given by $P_{ij} = xw_{ij}$, with $x = 2.59 \times 10–4$

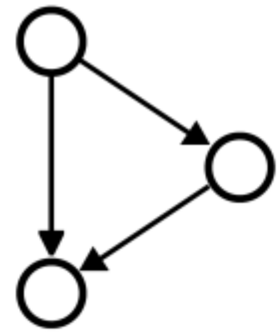


PNAS

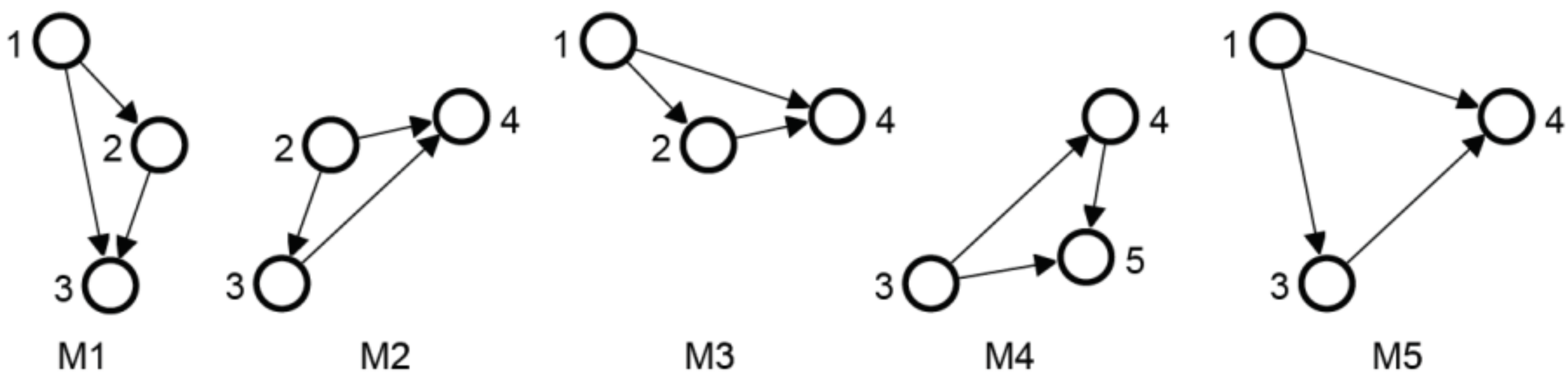
Resolving local structure: network motifs



graph



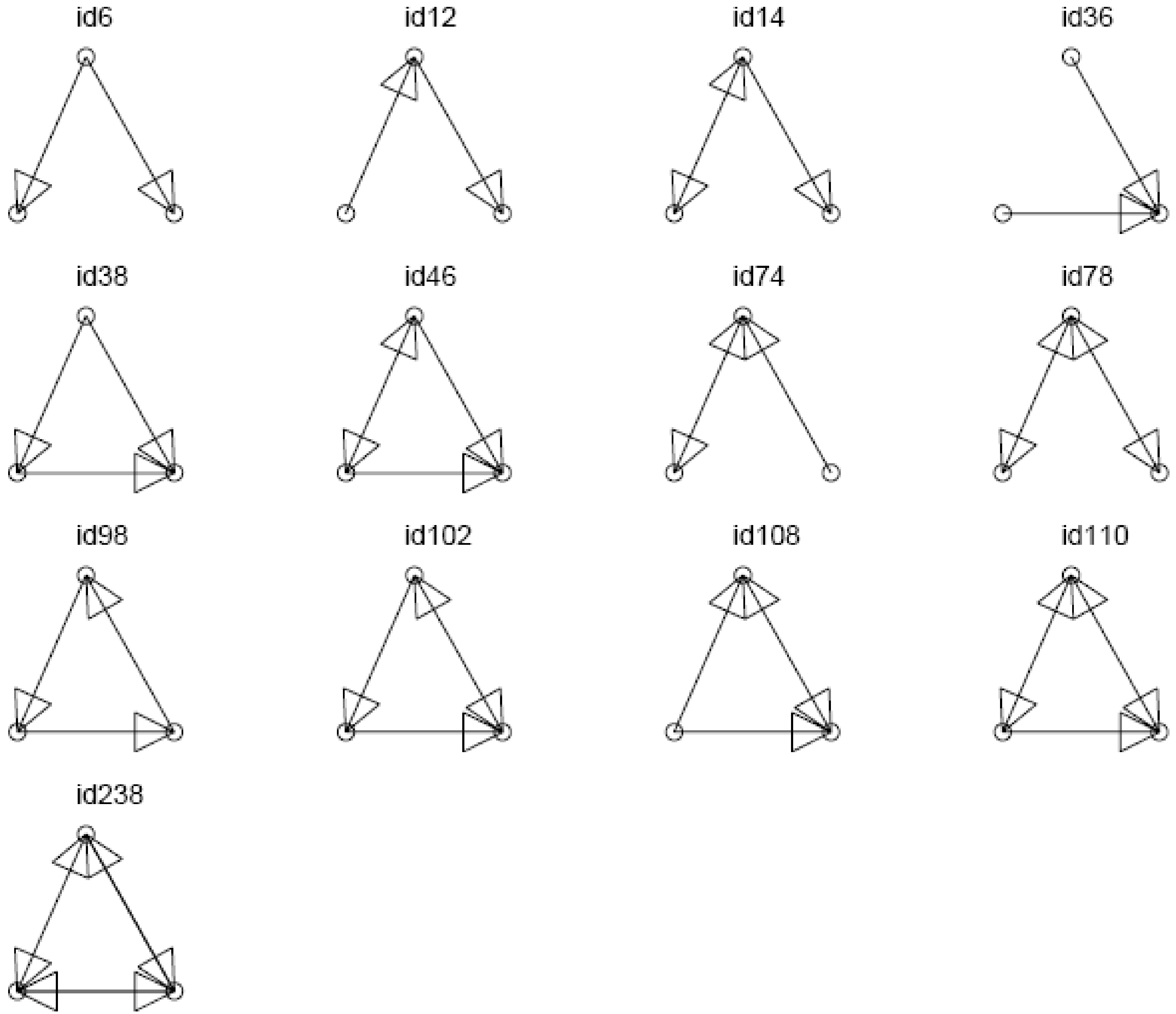
motif to be found



motif matches in the target graph

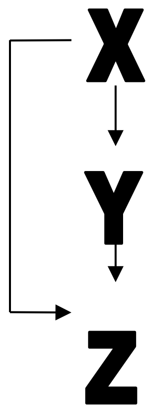
http://mavisto.ipk-gatersleben.de/frequency_concepts.html

All 3 node motifs

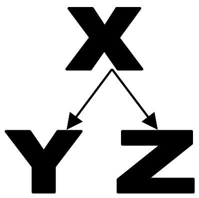


Examples of network motifs (3 nodes)

- Feed forward loop
  - Found in neural networks
  - Seems to be used to neutralize "biological noise"

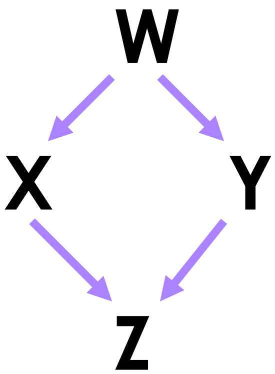


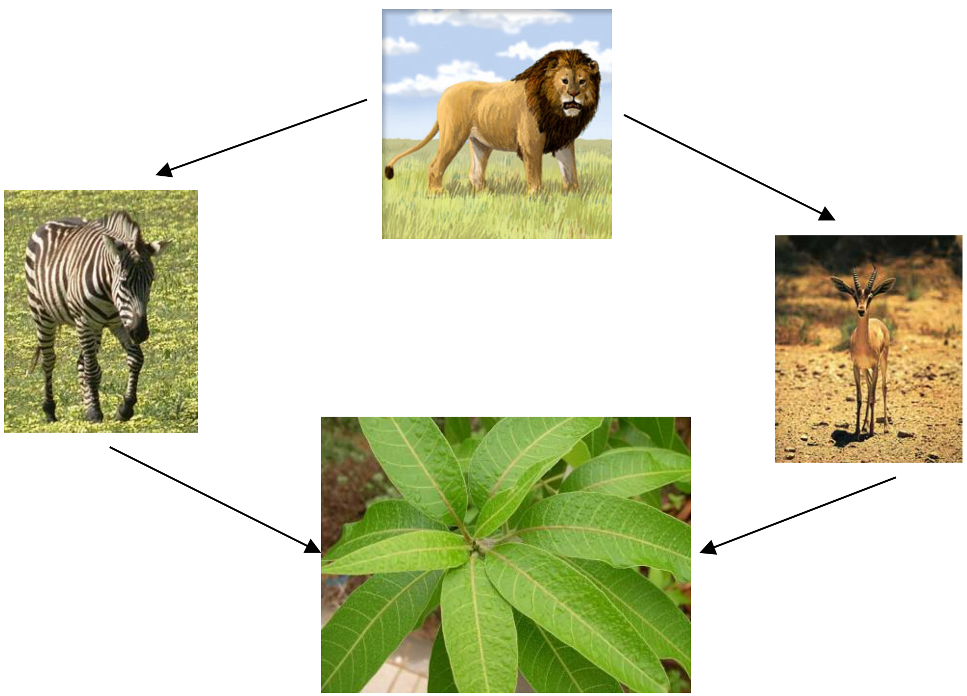
- Single-Input Module
  - e.g. gene control networks



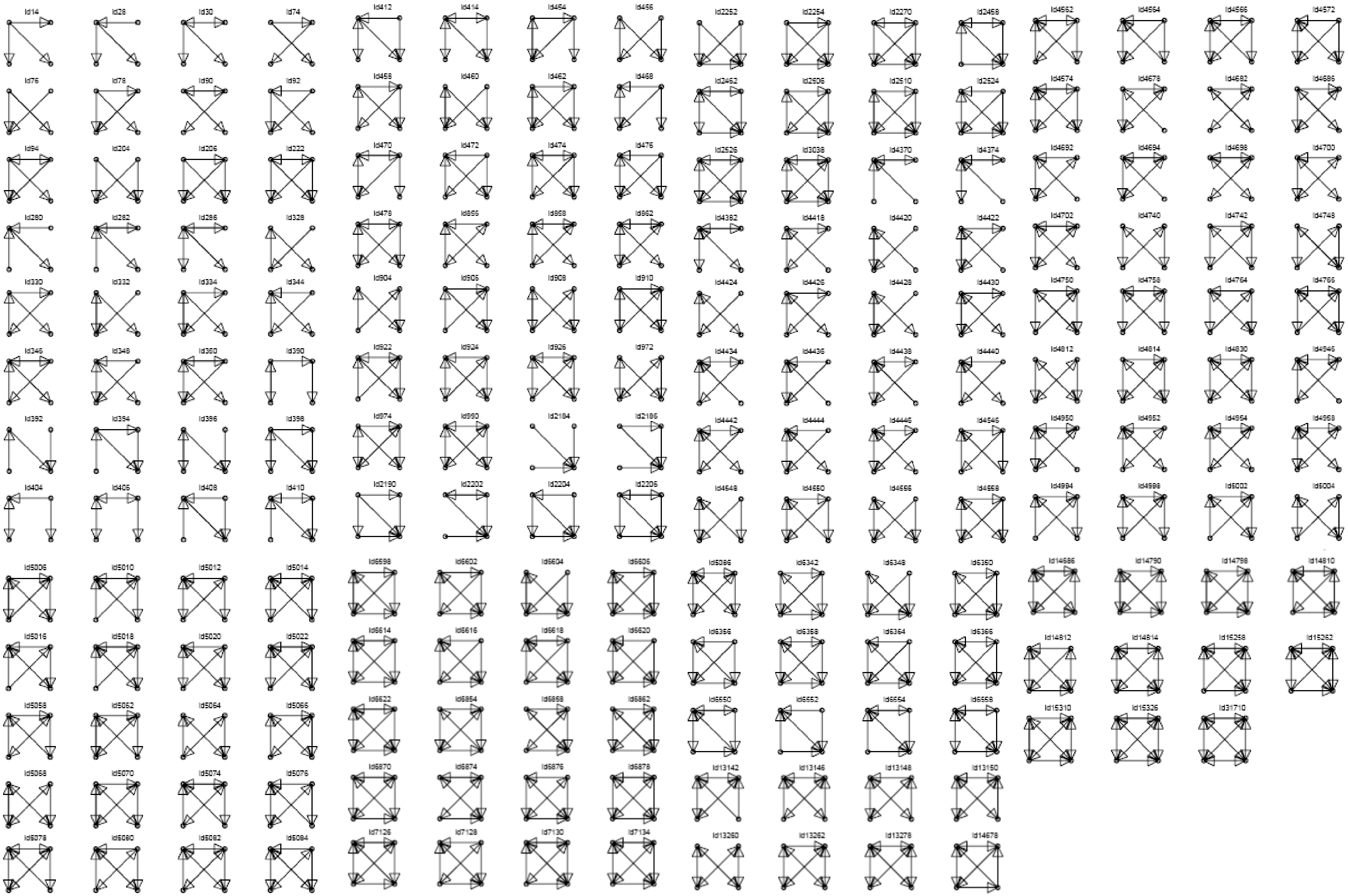
Examples of network motifs (4 nodes)

- Parallel paths
  - Found in neural networks
  - Food webs

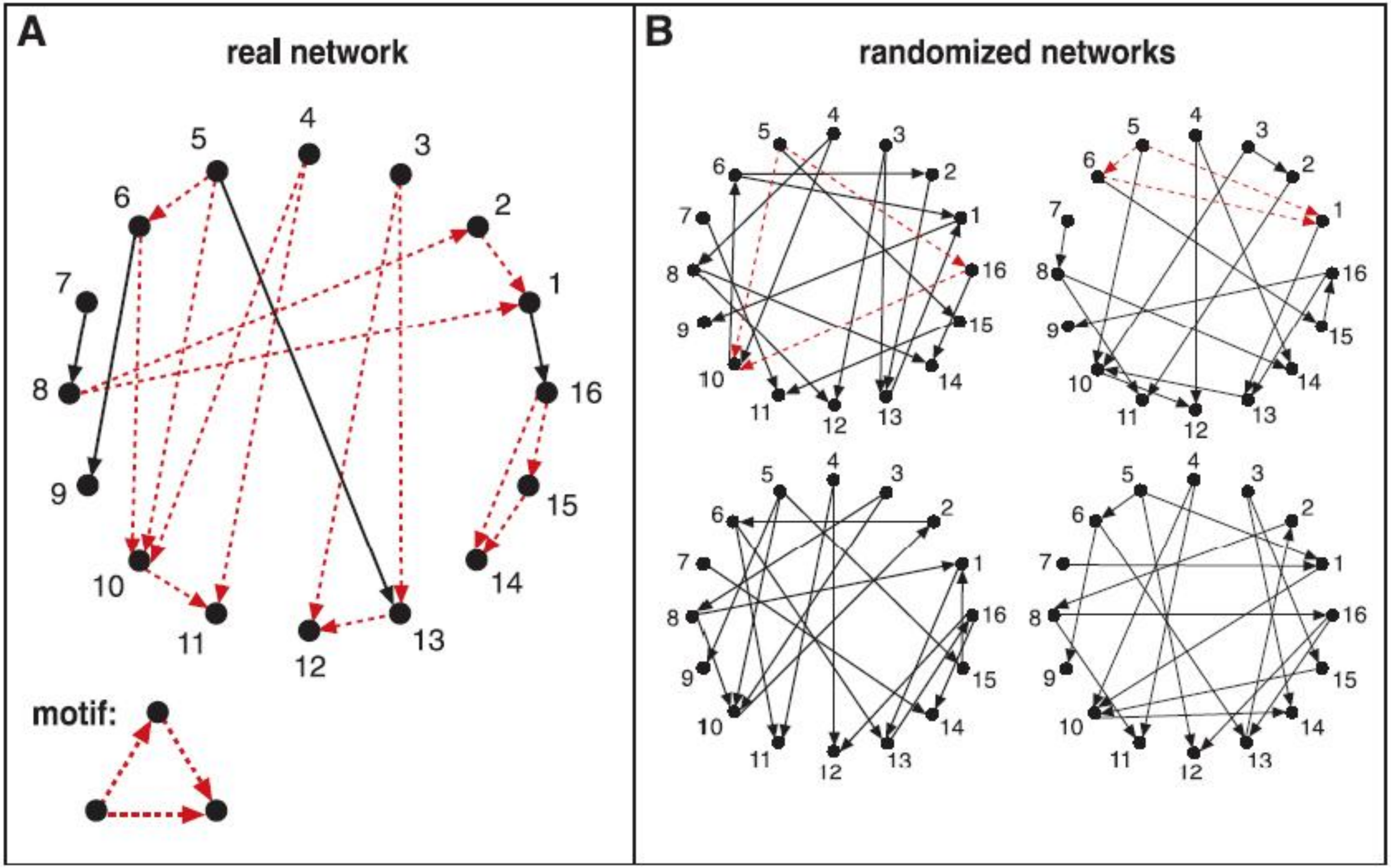




4 node subgraphs (computational expense increases with the size of the graph!)



Compare to "equivalent" random graph



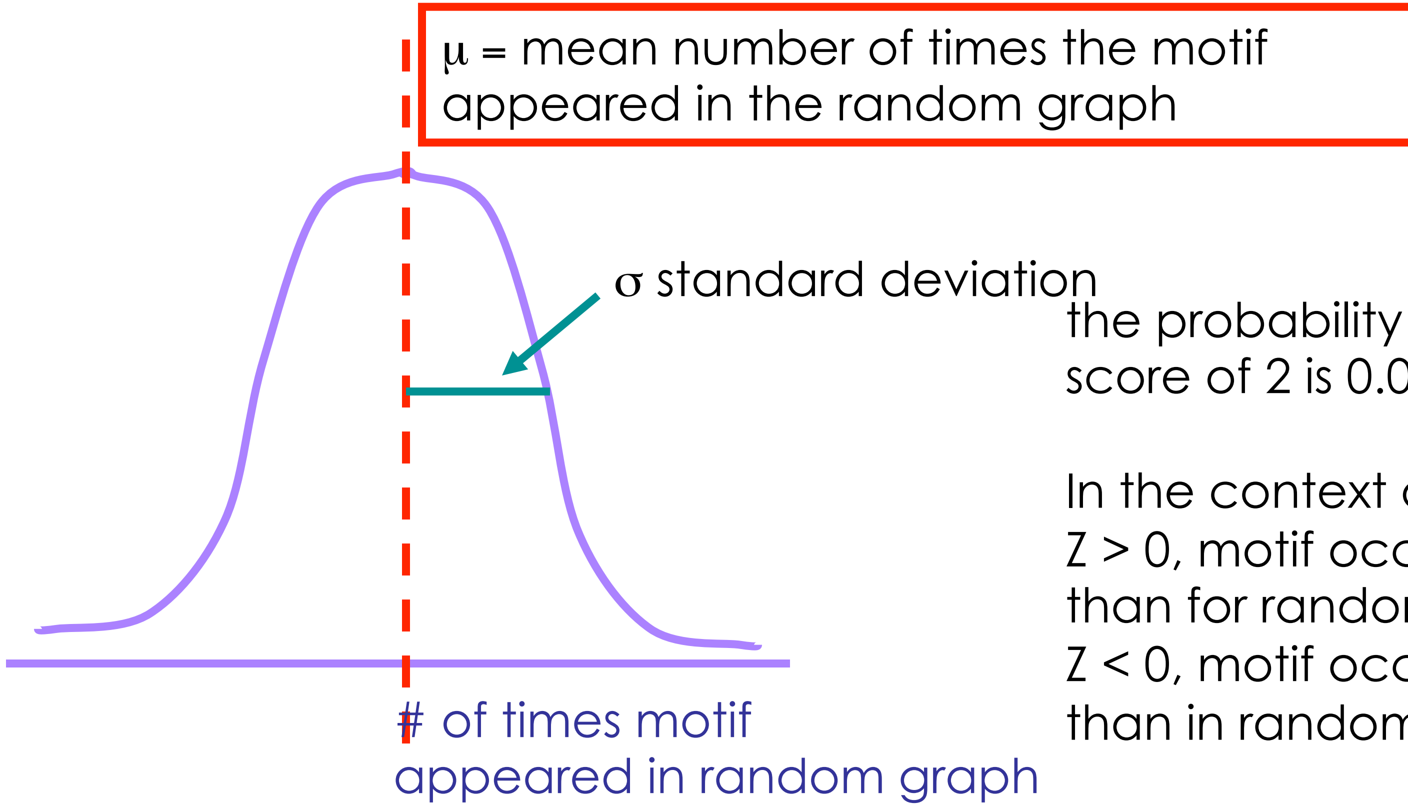
Milo et al., Network motifs: Simple building blocks of complex networks, Science 298:824-827, 2002

Network motif detection

- Some motifs will occur more often in real world networks than random networks
- Technique:
  - construct many random graphs with the same number of nodes and edges (same node degree distribution?)
  - count the number of motifs in those graphs
  - calculate the Z score: the probability that the given number of motifs in the real world network could have occurred by chance
- Software available:
  - http://www.weizmann.ac.il/mcb/UriAlon/ (the original)
  - http://theinf1.informatik.uni-jena.de/~wernicke/motifs/index.html

  (faster and more user friendly)

What the Z score means



$$z_x = \frac{x - \mu_x}{\sigma_x}$$

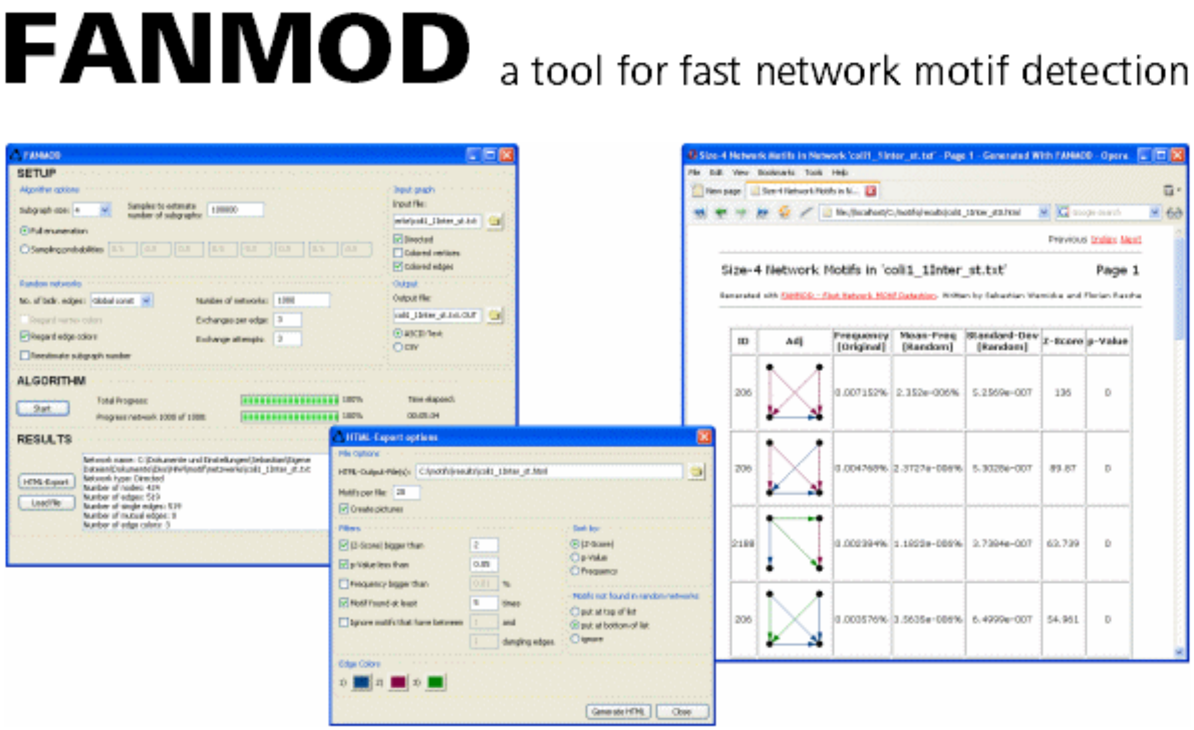
the probability observing a Z score of 2 is 0.02275

In the context of motifs:
Z > 0, motif occurs more often than for random graphs
Z < 0, motif occurs less often than in random graphs

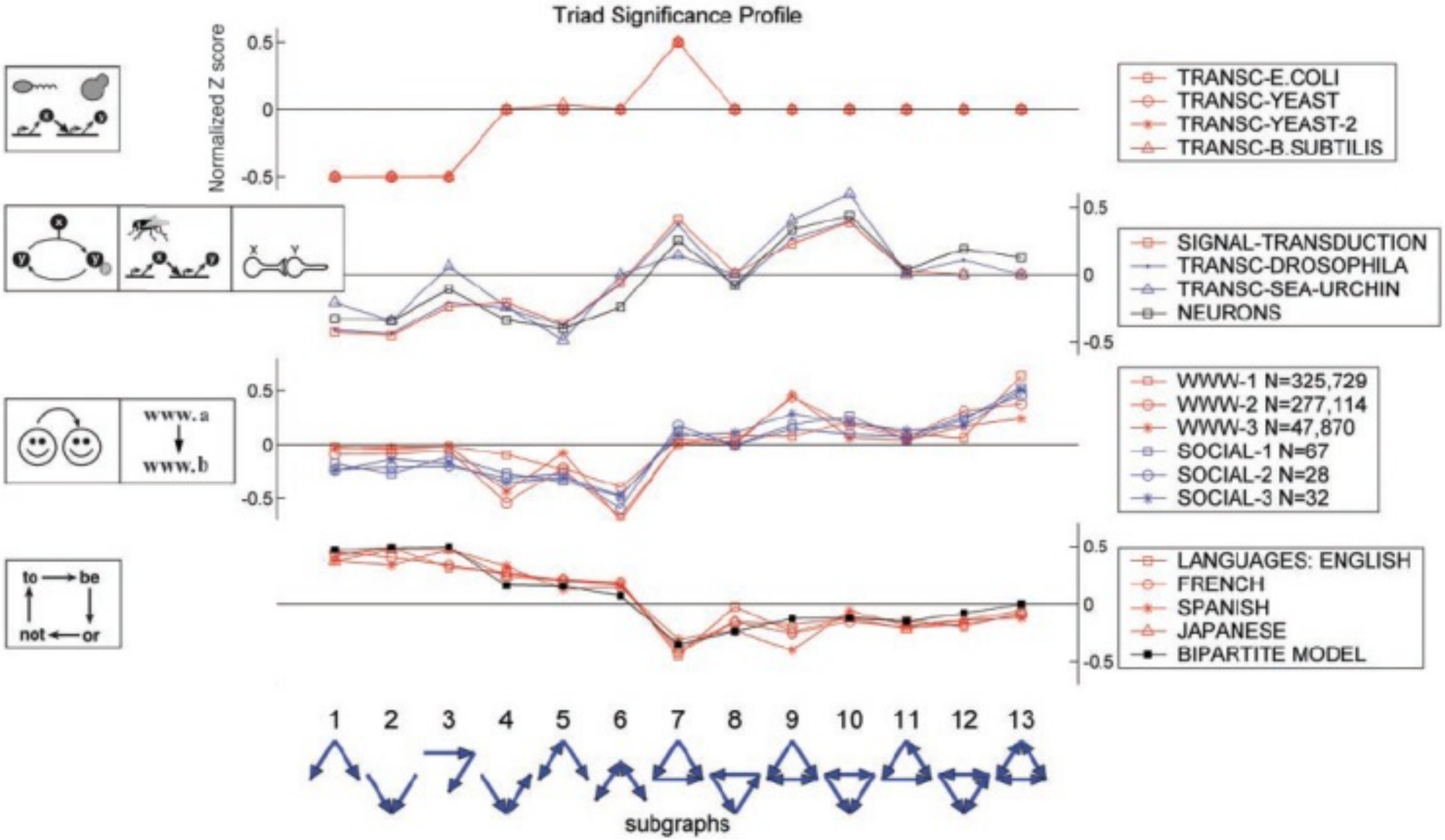
|Z| > 1.65, only a 5% chance of random occurence

software: FANMOD (also igraph)

- http://theinf1.informatik.uni-jena.de/~wernicke/motifs/index.html



Superfamilies of networks

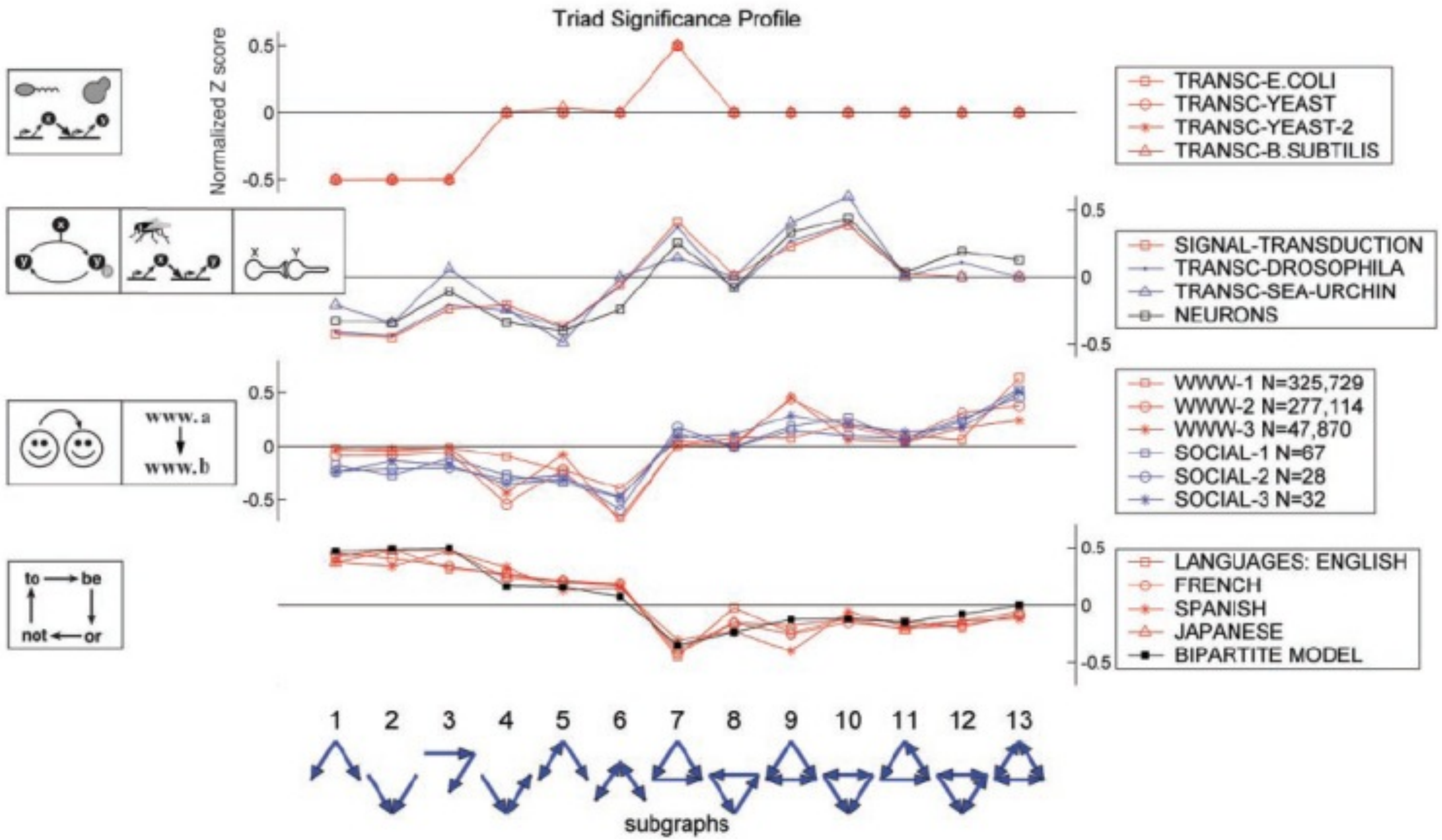


source: Milo et al., Superfamilies of Evolved and Designed Networks, Science 303:1538-1542, 2004

Quiz Q:

- Based on their triad census profiles, which two kinds of networks exhibit similar structure?

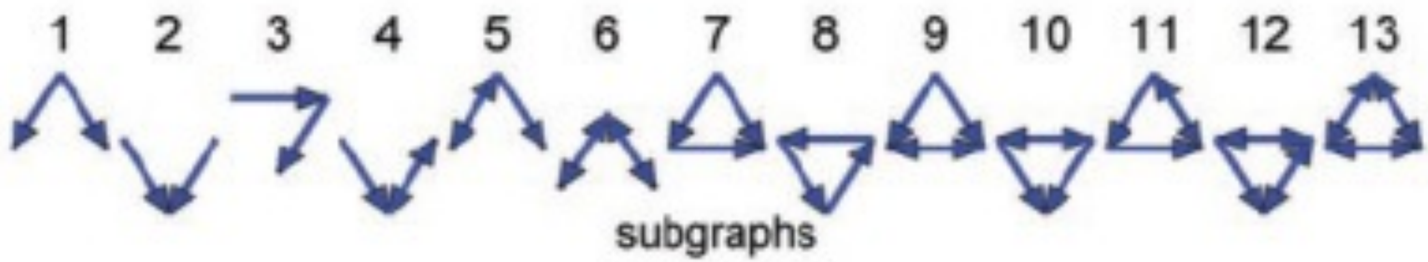
Superfamilies of networks



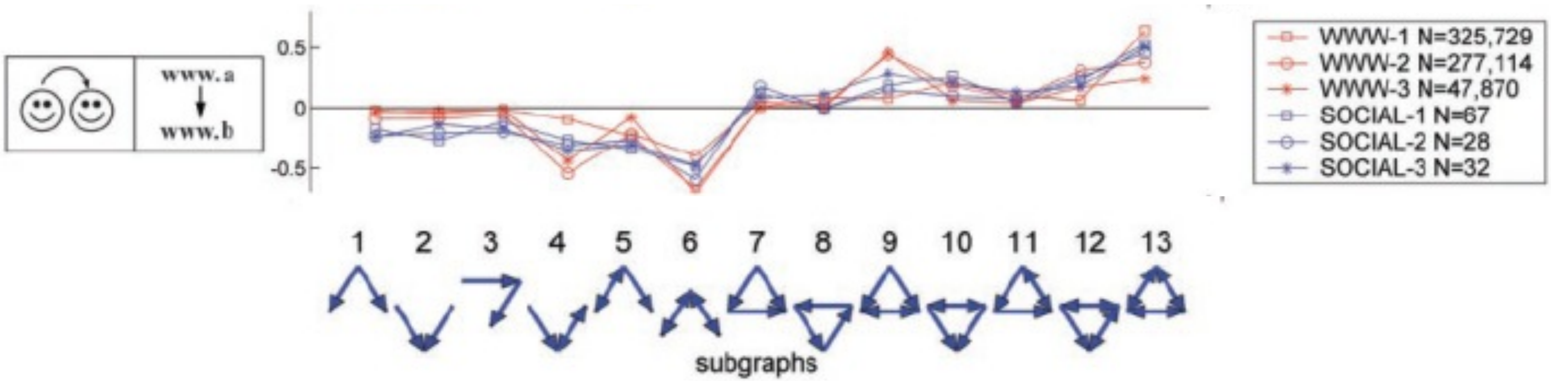
source: Milo et al., Superfamilies of Evolved and Designed Networks, Science 303:1538-1542, 2004

Quiz Q:

- Which of the following triads is underrepresented in social networks?



Superfamilies of networks



source: Milo et al., Superfamilies of Evolved and Designed Networks, Science 303:1538-1542, 2004

Motifs: recap

- Given a particular structure, search for it in the network, e.g. complete triads
- advantage: motifs can correspond to particular functions, e.g. in biological networks
- disadvantage: don't know if motif is part of a larger cohesive community