CIFF
Centro Internacior
de Formación Fina

CIFF Trustees:

Santander

Universidad de Alcalá

Social Network Analysis

Lecture 4: Small Worlds

David Martín-Corral Calvo
dmartincc@smartvel.com

Abril / Mayo

MASTER EN BA & BD

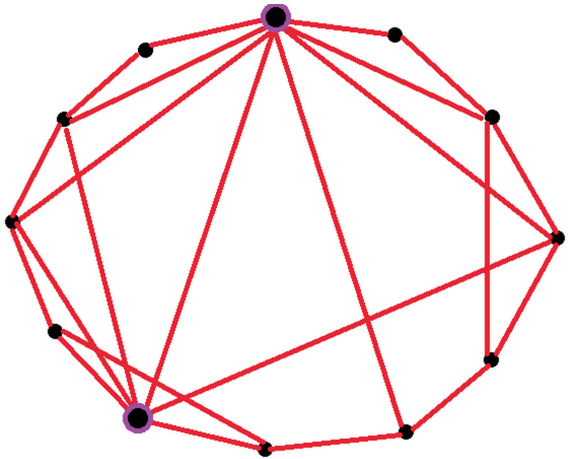
Recap:

- Community structure is a way of ‘x-raying’ the network, finding out what it’s made of.
- You can look for a specific structure:
  k-cliques, k-cores, etc.
- But most popular is to discover the “natural” community boundaries.

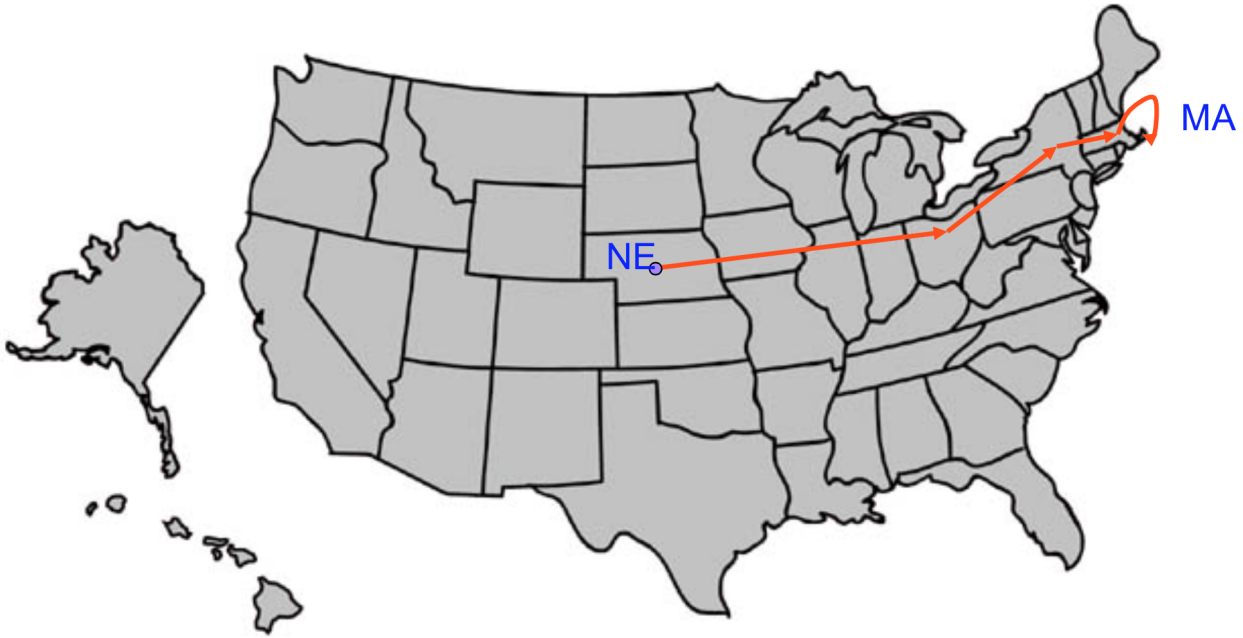
Outline:

- Milgram’s experiment
- Small worlds
- Motifs

A **small-world network** is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops or steps.



Small world phenomenon:
Milgram's experiment



Instructions:
Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target.

Outcome:

20% of initiated chains reached target
average chain length = 6.5

- "Six degrees of separation"

Milgram's experiment repeated:

email experiment
Dodds, Muhamad, Watts,
Science 301, (2003)
(optional reading)

- 18 targets
- 13 different countries

- 60,000+ participants
- 24,163 message chains
- 384 reached their targets
- average path length 4.0

Source: NASA, U.S. Government; http://visibleearth.nasa.gov/view_rec.php?id=2429

Interpreting Milgram's experiment

- Is 6 is a ***surprising*** number?
  - In the 1960s? Today? Why?

- Pool and Kochen in (1978 established that the average person has between 500 and 1500 acquaintances)

Quiz

- Ignore for the time being the fact that many of your friends' friends are your friends as well. If everyone has 500 friends, the average person would have how many friends of friends?

Quiz

- With an average degree of 500, a node in a random network would have this many friends-of-friends-of-friends (3rd degree neighbors):

Interpreting Milgram's experiment

- Is 6 is a ***surprising*** number?
  - In the 1960s? Today? Why?
- If social networks were random… ?
  - Pool and Kochen (1978) - ~500-1500 acquaintances/person
  - ~ 500 choices 1st link
  - ~ 500^2 = 250,000 potential 2nd degree neighbors
  - ~ 500^3 = 125,000,000 potential 3rd degree neighbors
- If networks are completely cliquish?
  - all my friends' friends are my friends
  - what would happen?

Quiz

- If the network were completely cliquish, that is all of your friends of friends were also directly your friends, what would be true:

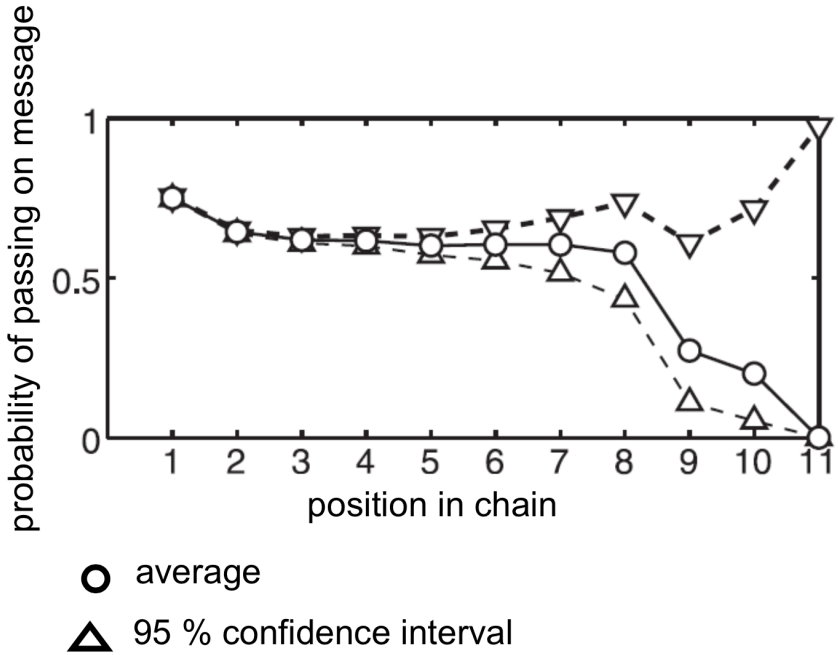
Complete cliqueness

- If all your friends of friends were also your friends, you would be part of an isolated clique.

Uncompleted chains and distance

- Is 6 an ***accurate*** number?
- What bias is introduced by uncompleted chains?
  - are longer or shorter chains more likely to be completed?

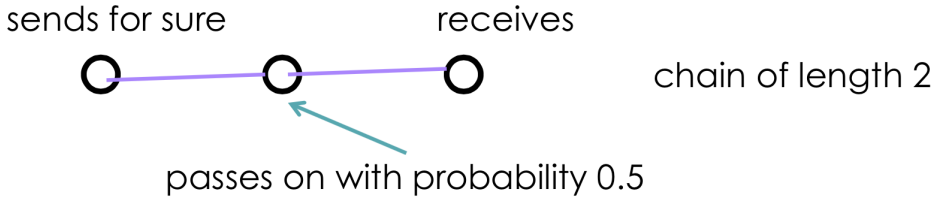
Attrition



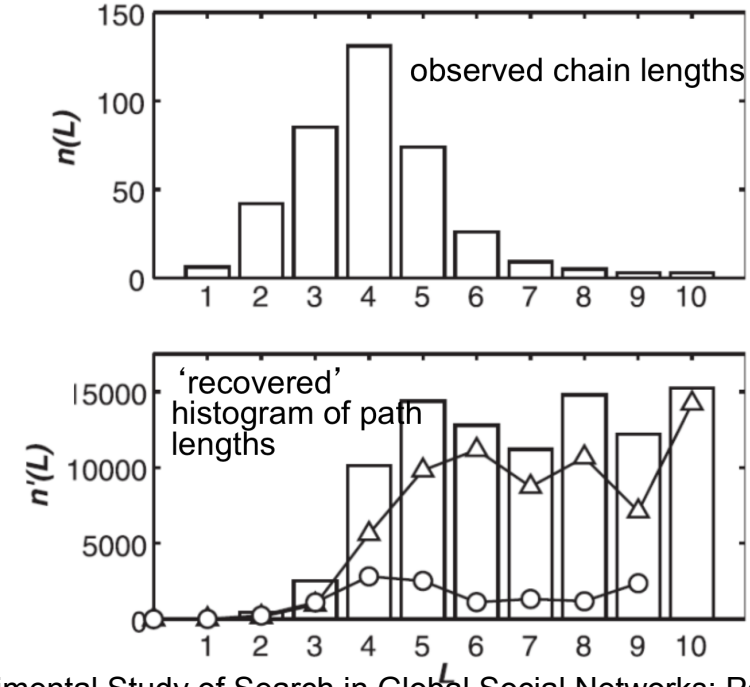
Source: An Experimental Study of Search in Global Social Networks: Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts (8 August 2003); Science 301 (5634), 827

Quiz

- if each intermediate person in the chain has 0.5 probability of passing the letter on, what is the likelihood of a chain being completed
  - of length 2?
  - of length 5?



Estimating the true distance



inter-country
intra-country

Source: An Experimental Study of Search in Global Social Networks: Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts (8 August 2003); Science 301 (5634), 827

Navigation and accuracy

- Is 6 an ***accurate*** number?

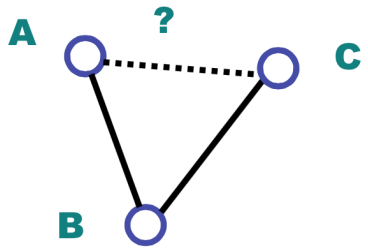
- Do people find the ***shortest*** paths?
  - Killworth, McCarty ,Bernard, & House (2005, optional):
  - less than optimal choice for next link in chain is made ½ of the time

Small world and networking

What does it mean to be 1, 2, 3 hops apart on Facebook, Twitter, LinkedIn, Google Plus?

Transitivy, triadic closure and 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



Global Clustering Coefficient

- Global clustering coefficient

  3 x number of triangles in the graph

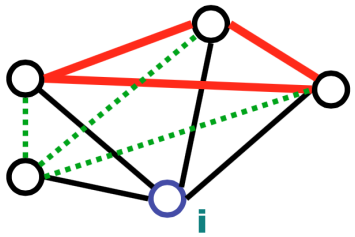
  number of connected triples of vertices

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

Local clustering coefficient

- Average over all n vertices

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

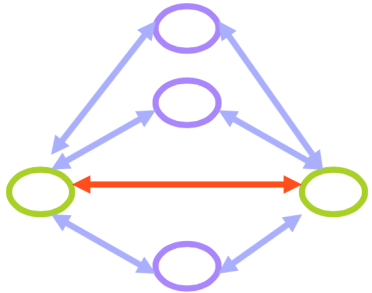


link present
link absent

$n_i = 4$
max number of connections:
4*3/2 = 6
3 connections present
$C_i = 3/6 = 0.5$

Embeddedness

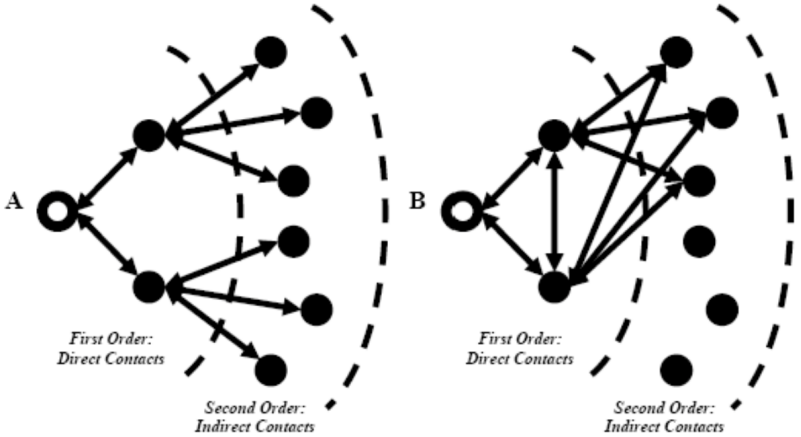
- embeddeness: number of common neighbors the two endpoints have



- neighborhood overlap:

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

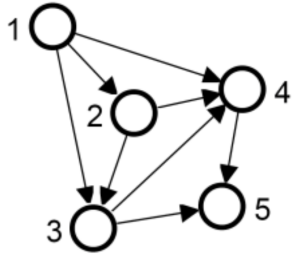
- snowball sampling:
  - will you reach more different kids by asking each kid to name their 2 best friends, or their 7th & 8th closest friend?



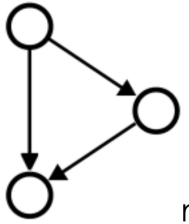
Is it good to be embedded?

- 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?

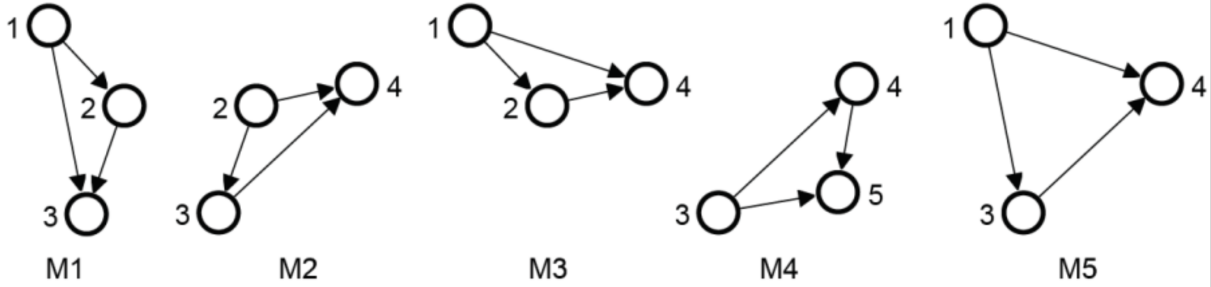
Resolving local structure: network motifs



graph

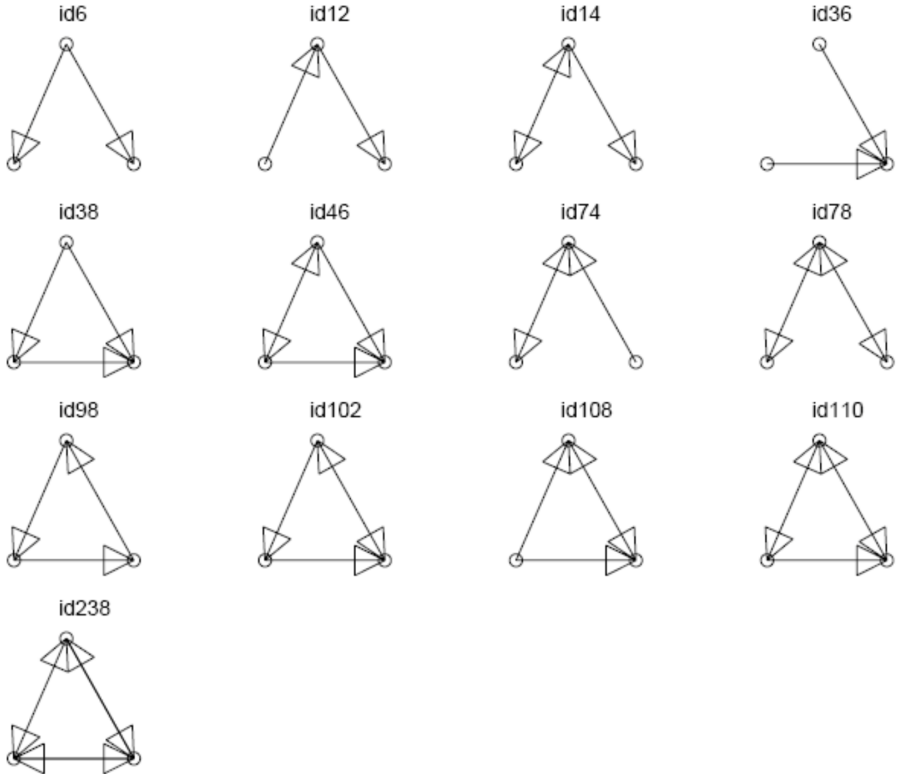


motif to be found



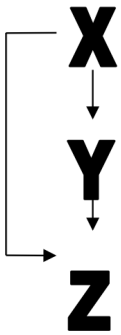
motif matches in the target graph

All 3 node motifs

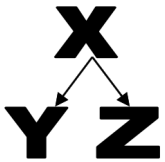


Examples of 3 nodes motifs

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



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



Recap:

- The world is small if you look at it as a network
- It has lots of interesting structures.
- Motifs: Given a particular structure, search for it in the network, e.g. complete triads

Activity: networkx

Python script

Reading list:

An Experimental Study of Search in Global Social Networks

Suggested links:

networkx documentation

CIFF

Centro Internacional
de Formación Financiera

CIFF Trustees:

Santander

Universidad de Alcalá