CISC 372
Network/Graph I

There's so much interest in social networks because:

each edge is created by a local decision

BUT global structure emerges from the totality of the local decisions

(and any individual can't control this global structure, so it provides an unfiltered view of groups)

A social network:

- a set of nodes, representing individuals
- a set of edges connecting (some) pairs of nodes, reflecting a relationship between the corresponding individuals

Edges can be directed or undirected

(most relationships are directed, but most social networks ignore this)

Edges can be typed (e.g. friend or colleague)

Edges can be signed (friends vs enemies)

Edges can represent: flow of information, flow of influence, dominance, tie signs

Real-world social networks

Sizes are not arbitrary

– Dunbar limit

- Powers of “almost 3”

(probably a property of language)

Influence in social networks often travels 3 steps, so you are influenced by people you don't even know!

Directly connected to a happy person: 15% more likely to be happy

2 steps away from a happy person: 10% more likely to be happy

3 steps away from a happy person: 6% more likely to be happy

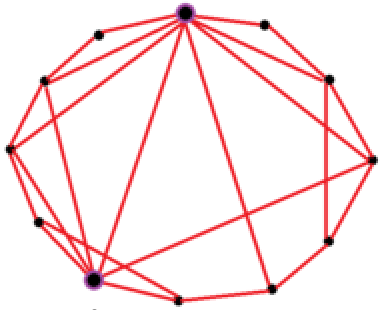
Directly connected to an unhappy person: 7% less likely to be happy

The same holds for:

* Obesity (friend increase your risk by 57%)
* Political views
* Weight gain
* Smoking
* Altruism ($1 produces $0.20 X 3 at the second level and $0.05 X 9 at the third)
* Suicide, depression, violence, back pain, getting flu shots, …

Real-world social networks

Small world property – the diameter of the graph is much smaller than expected



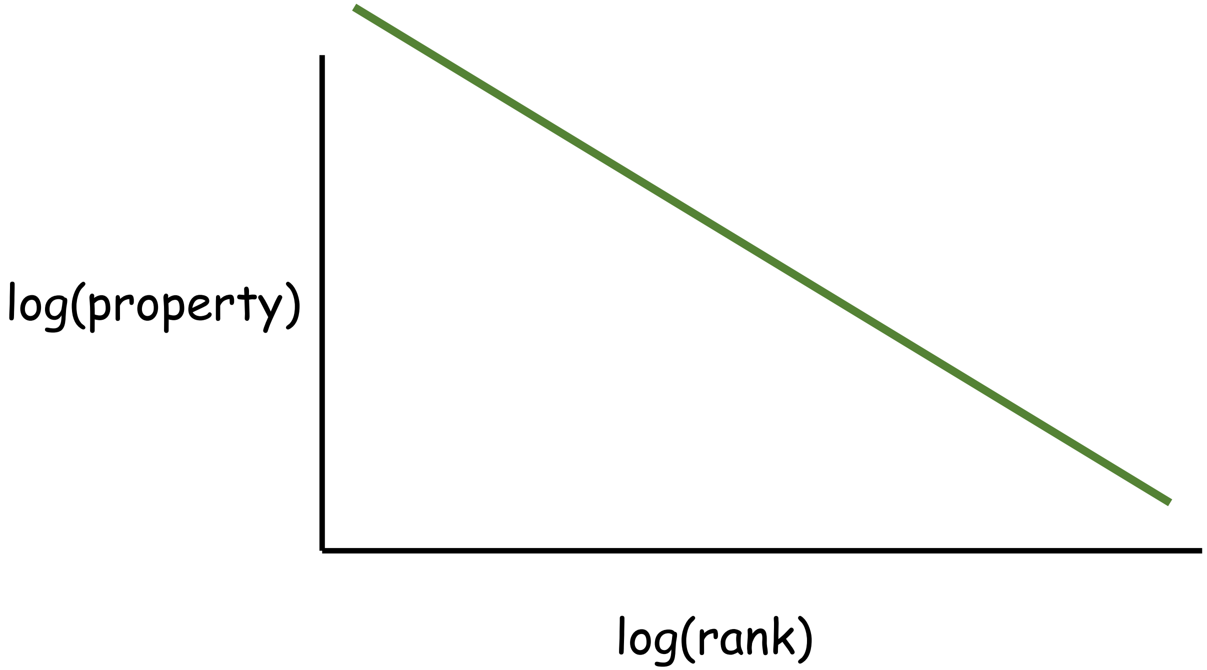
6 degrees of separation

all people are six, or fewer, social connections away from each other

Therefore no real concept of clusters

Power laws:

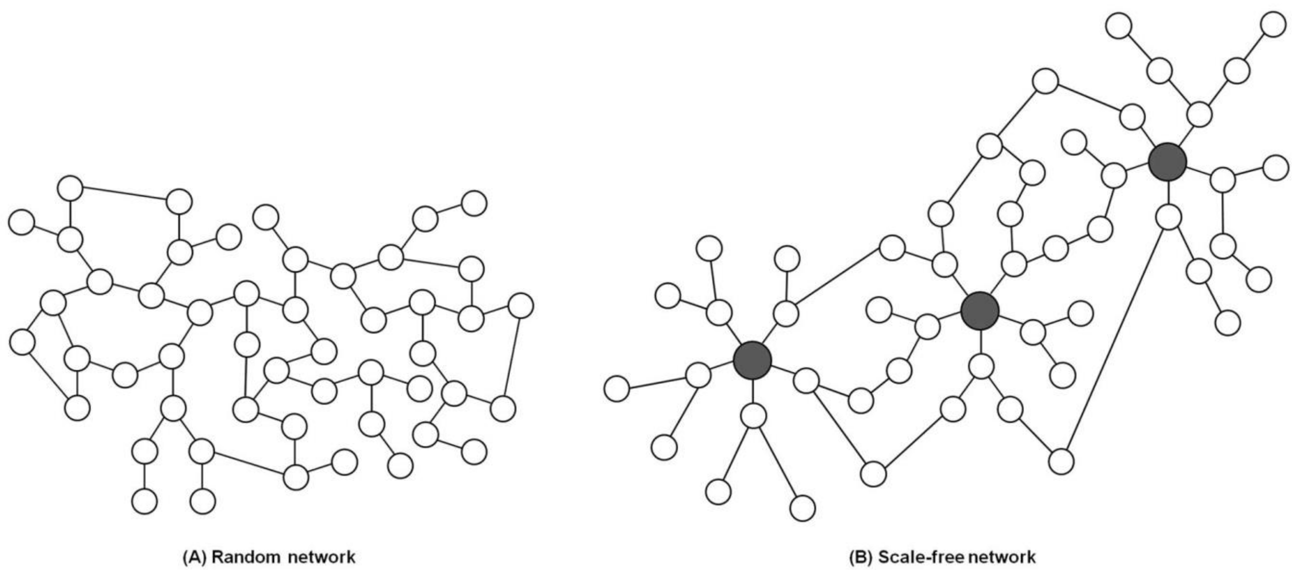
Many properties have distributions like this:



and so,

scale invariance – things look the same in any piece of the graph

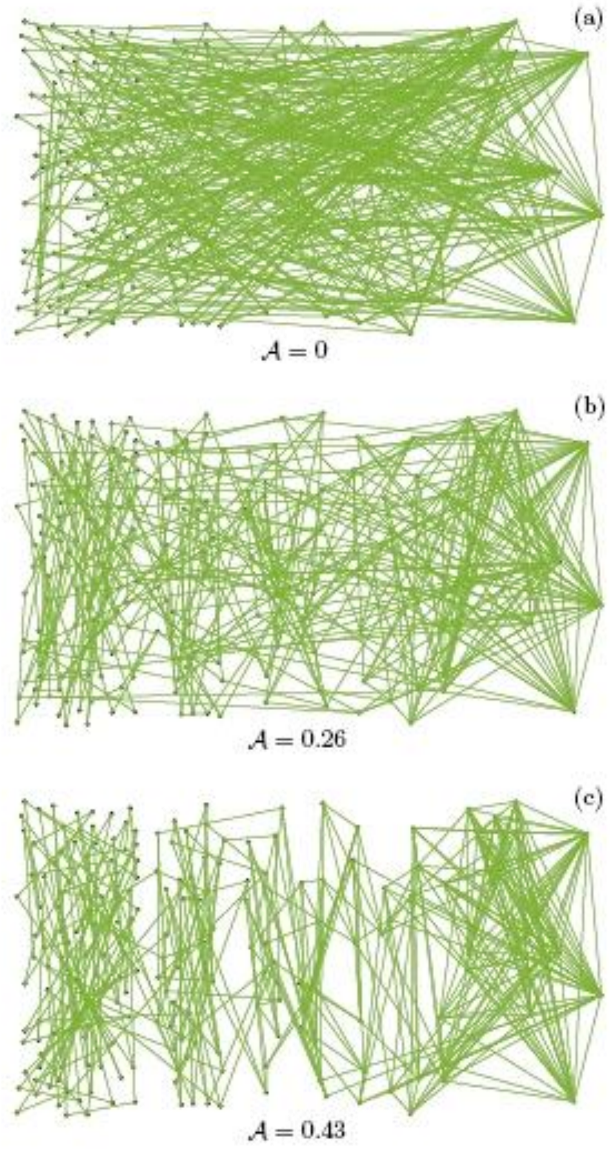
averages are meaningless



Other properties:

Assortativity

- i.e. correlation between two nodes
- Many measurements

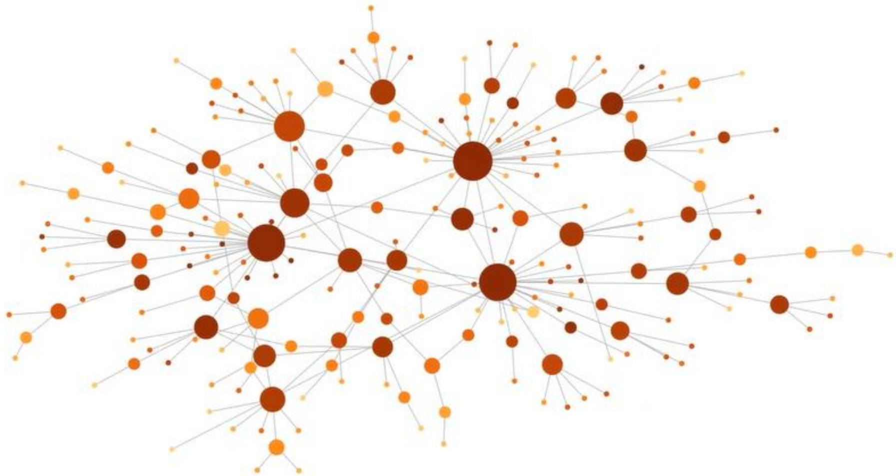


Attachment mechanisms

As real networks were examined more deeply, models of attachment have become more complex:

random graphs

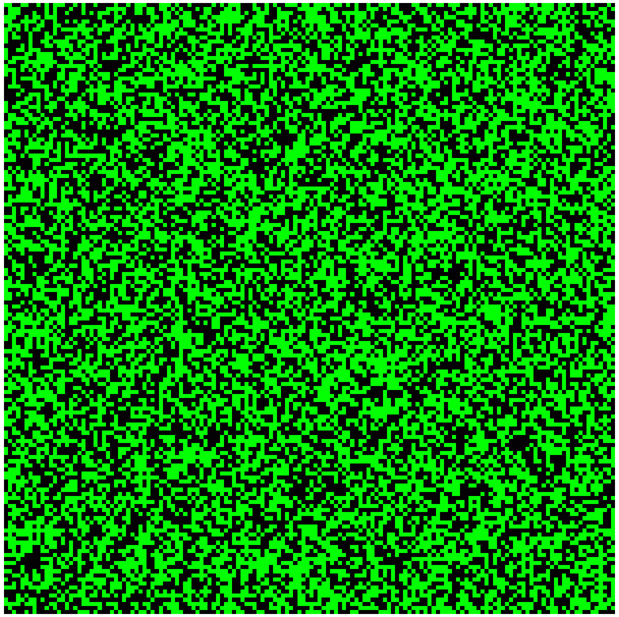
preferential attachment



https://vimeo.com/53071346

Forest fire attachment (hairball and whiskers)

(forest file model)



Social network analysis

What structures are present in the network?

Which nodes are most important (as controllers; as influencers; as connectors of others)?

Which subgraphs are most important?

If one new edge came into existence, which nodes is it most likely to connect (a.k.a. “this might be someone you know”)?

Answers to these questions can usually be computed directly on the graph, BUT …

- the time complexity is often cubic in the size of the graph

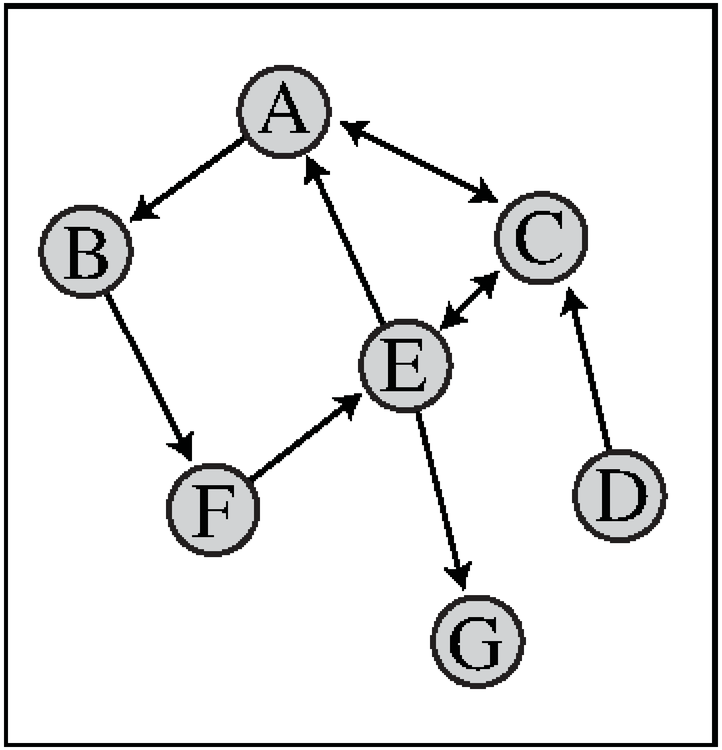
- changing even a single node or edge can require recomputing from scratch

So it’s usual to embed the graph in a geometry:

each node is mapped to a position in a k-dimensional space ($k \ll n$)

in such a way that closeness in the space represents closeness in the graph.

If we start from an adjacency matrix,



	A	B	C	D	E	F	G
A	0	1	1	0	0	0	0
B	0	0	0	0	0	1	0
C	1	0	0	0	1	0	0
D	0	0	1	0	0	0	0
E	1	0	1	0	0	0	1
F	0	0	0	0	1	0	0
G	0	0	0	0	0	0	0

Subgraphs?

In human networks, triangles are especially important – communication and resilience

For n nodes, about $n^{1.8}$ triangles

1000 nodes -> 251,189 triangles

Build a fresh graph, with the same nodes, keeping edges that participate in many triangles.

Embed this graph and find the structure of the key players (Simmelian backbone)

Edge prediction?

Predict the pair of nodes that is closest in the embedding, but not already connected

e.g. Facebook, Linkedin

Summary:

Social networks are useful data because their properties are emergent

Social networks have unexpected properties that have increased our understanding of ourselves and our social structures

The graphs that represent social networks require more complex analytic methods that are still being actively developed.