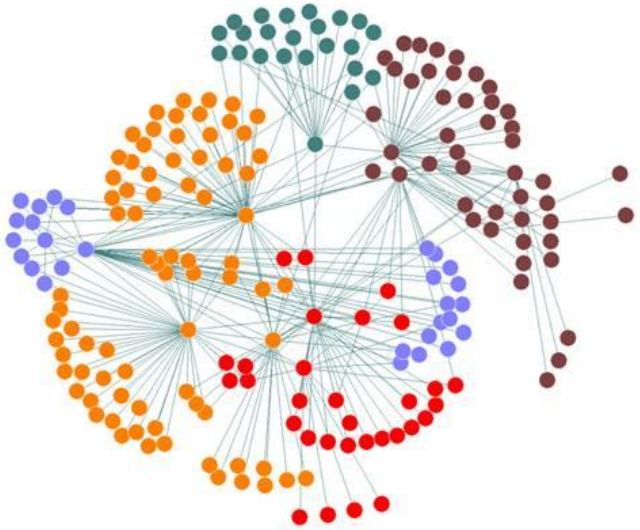
TEL AVIV UNIVERSITY

Algorithms and Applications in Social Networks



2019/2020, Semester B

Slava Novgorodov

Lesson #1

- Administrative questions
- Course overview
- Introduction to Social Networks
- Basic definitions
- Network properties

Administrative questions

- Course format:
  - Lecture (2h) + Recitation (1h) every week (by Slava)
  - 3 Homework tasks during the semester
    - Submission in pairs
    - Theoretical + Practical (Python) questions
  - Final exam (format will be discussed later)
  - Final grade = 85% Exam + 15% HW

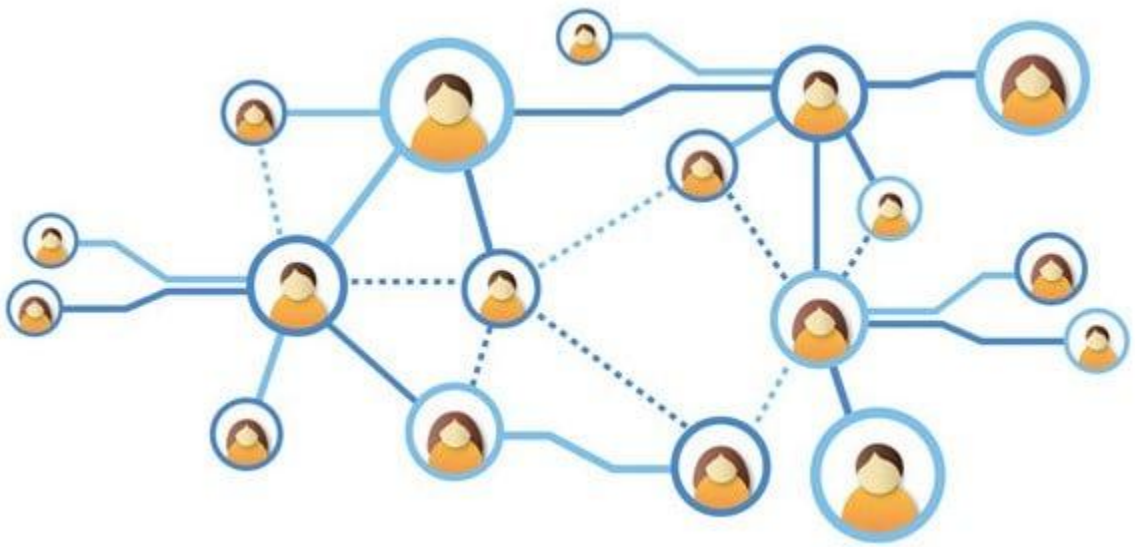
  - Office hours – Sunday (schedule in advance)
  - Course website: http://slavanov.com/teaching/sn1920b/
  - Email: slavanov@**post**.tau.ac.il (not mail.tau.ac.il !)

Related material

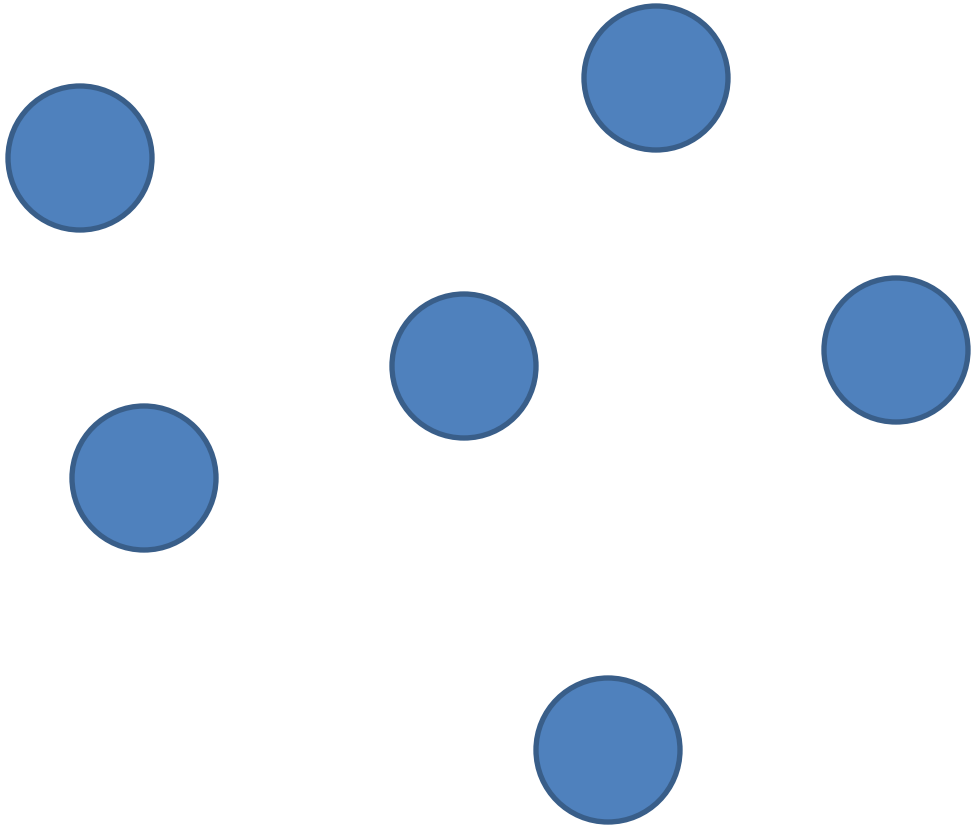
- Books:
  - **Newman** “Networks: An Introduction”
  - **Jackson** “Social and Economic Networks”
  - **Easley** & **Kleinberg** “Networks, Crowds, and Markets: Reasoning About a Highly Connected World” http://cs.cornell.edu/home/kleinber/networks-book/
  - **Wasserman** & **Faust** “Social Network Analysis. Methods and Applications.”
- Related courses:
  - CS224W (Stanford) – Analysis of Networks https://web.stanford.edu/class/cs224w/
  - Social and Economics networks (online course) https://www.youtube.com/channel/UCCnG8fKY45aH73ahmGK2xcg
  - High School of Economics – Social Networks http://leonidzhukov.net/hse/2014/socialnetworks/

Social Networks

- **Social Network** - a structure of social actors (individuals or organizations) and social interactions between the actors

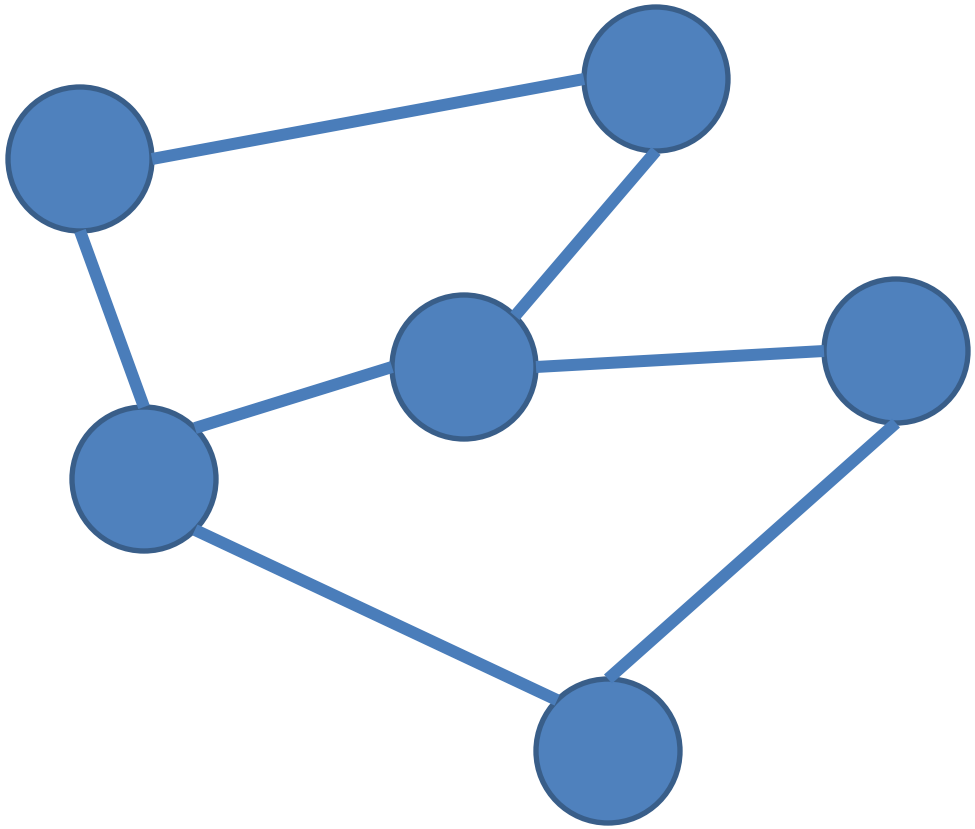


Social Networks



Social actors

Social Networks



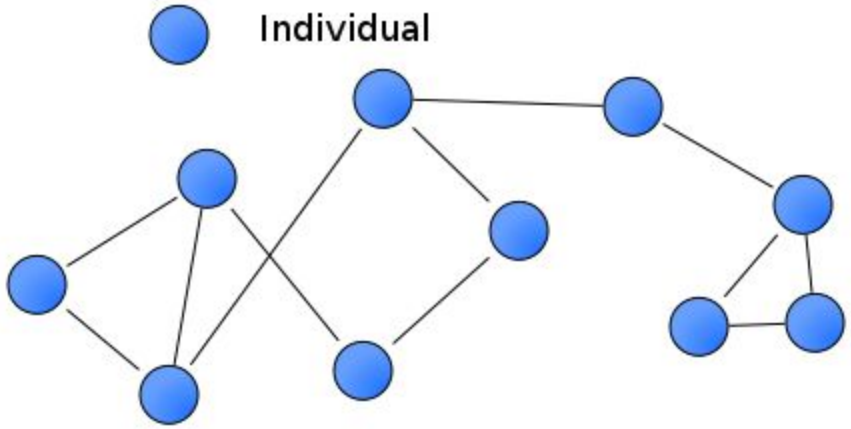
Social actors and **interactions**

Social Networks

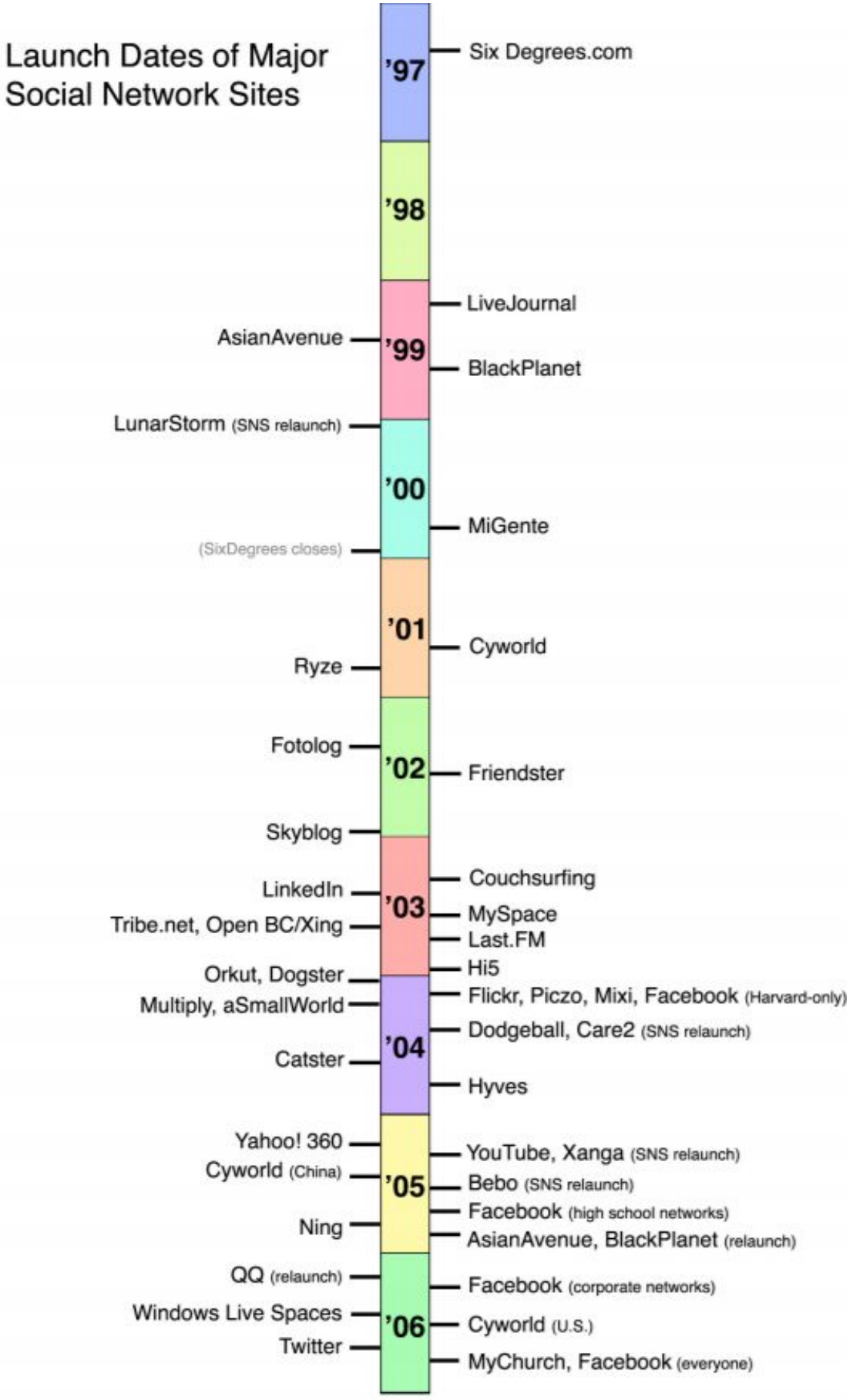
- **Interdisciplinary** field, studied in:
  - Sociology
  - Social psychology
  - Economics
  - Statistics
  - Mathematics (Graph Theory)
  - Computer Science (this course)

Social Networks

- The research around Social Networks started at the beginning of 1930s (first sociograms)



- Mathematical formulation – 1950s
- 1980s and later – growth in number of social network research and researchers
- Late 1990s until now – **online** social networks


Launch Dates of Major Social Network Sites
'97
Six Degrees.com
'98
'99
LiveJournal
AsianAvenue
BlackPlanet
LunarStorm (SNS relaunch)
'00
MiGente
(SixDegrees closes)
'01
Cyworld
Ryze
Fotolog
'02
Friendster
Skyblog
'03
Couchsurfing
LinkedIn
MySpace
Tribe.net, Open BC/Xing
Last.FM
Hi5
Orkut, Dogster
Flickr, Piczo, Mixi, Facebook (Harvard-only)
Multiply, aSmallWorld
'04
Dodgeball, Care2 (SNS relaunch)
Catster
Hyves
Yahoo! 360
YouTube, Xanga (SNS relaunch)
Cyworld (China)
'05
Bebo (SNS relaunch)
Facebook (high school networks)
Ning
AsianAvenue, BlackPlanet (relaunch)
QQ (relaunch)
Facebook (corporate networks)
Windows Live Spaces
'06
Cyworld (U.S.)
Twitter
MyChurch, Facebook (everyone)

Research clusters

- Communications
- Complex networks
- Criminal networks
- Spread of innovations
- Demography
- Health care
- Language and linguistics
- Social media
- …

What can be presented as SN?

- Friendship and other social relationships
- Corporative structures (internal/external)
- Trade relationships (individuals/companies)
- Political alliances
- Sharing of information
- Criminal organizations structures
- …

Three aspects

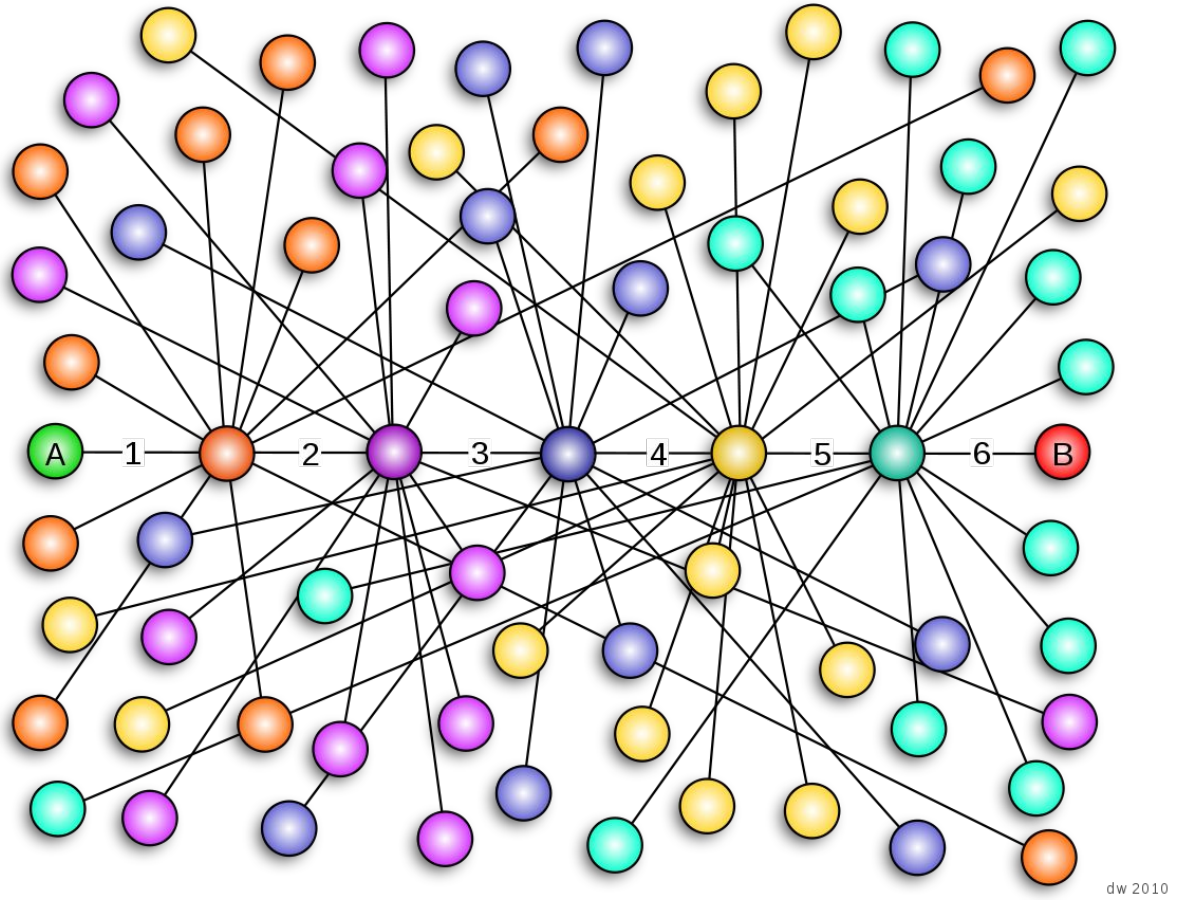
- Theory
  - Network formation, dynamics…
  - Influence detection
  - Communities
- Experimental studies
  - Observe patterns
  - Test theories
- Methodology
  - How to analyze networks?

Applications in Social Networks

Facebook Social Graph

2 billion monthly users - which is ~70% of the 2.8 billion Internet users living outside of China/Russia (they use a different social networking system).

6 degrees of separation



6 degrees of separation

The Small World experiment:

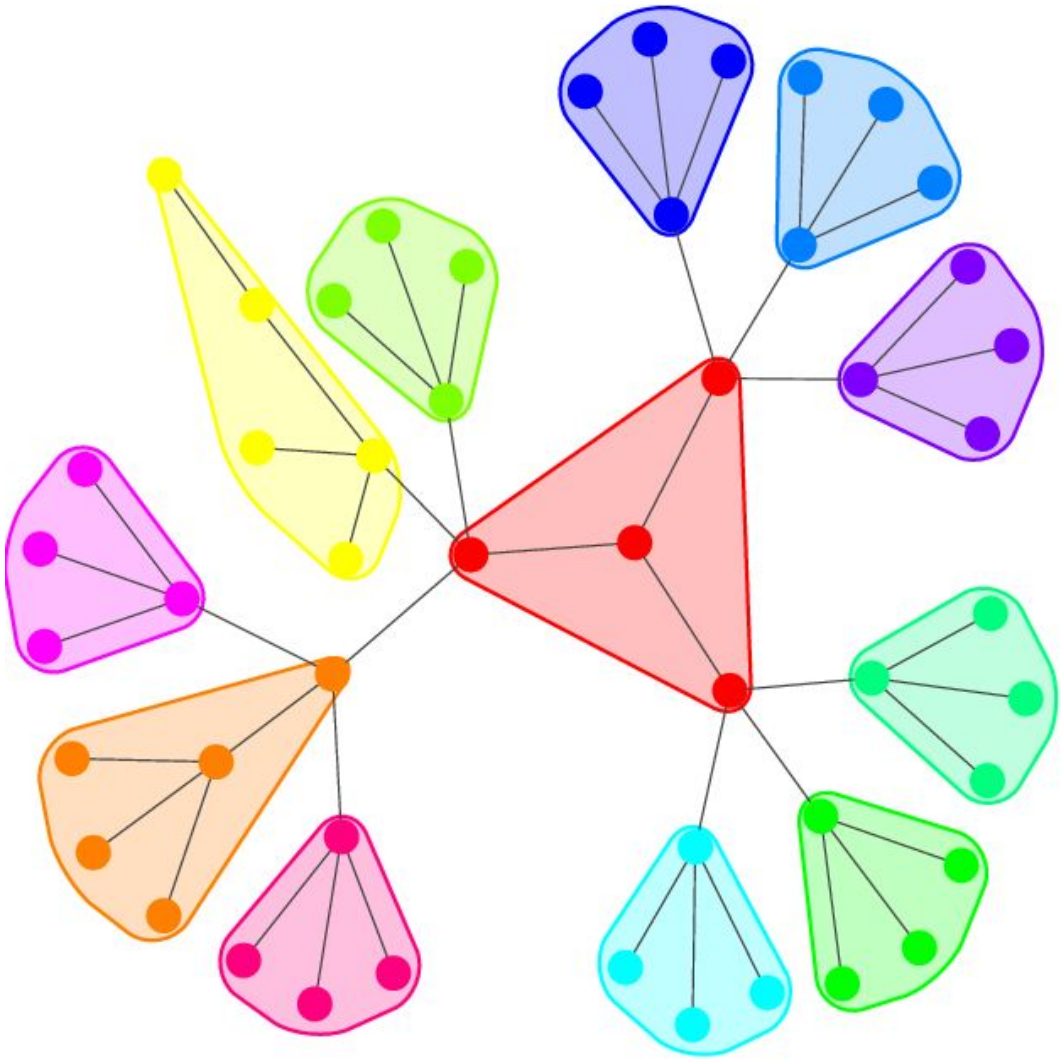
Model the population as a social network and attempt to find the average path length between any two nodes.

1. Select individuals in two far (socially and geographically) points – Omaha, Nebraska and Boston, Massachusetts
2. The individual in Omaha received a letter he/she needs to pass to an individual in Boston. If they know each other, great. Otherwise, the letter should be sent to a friend who may know the destination individual.

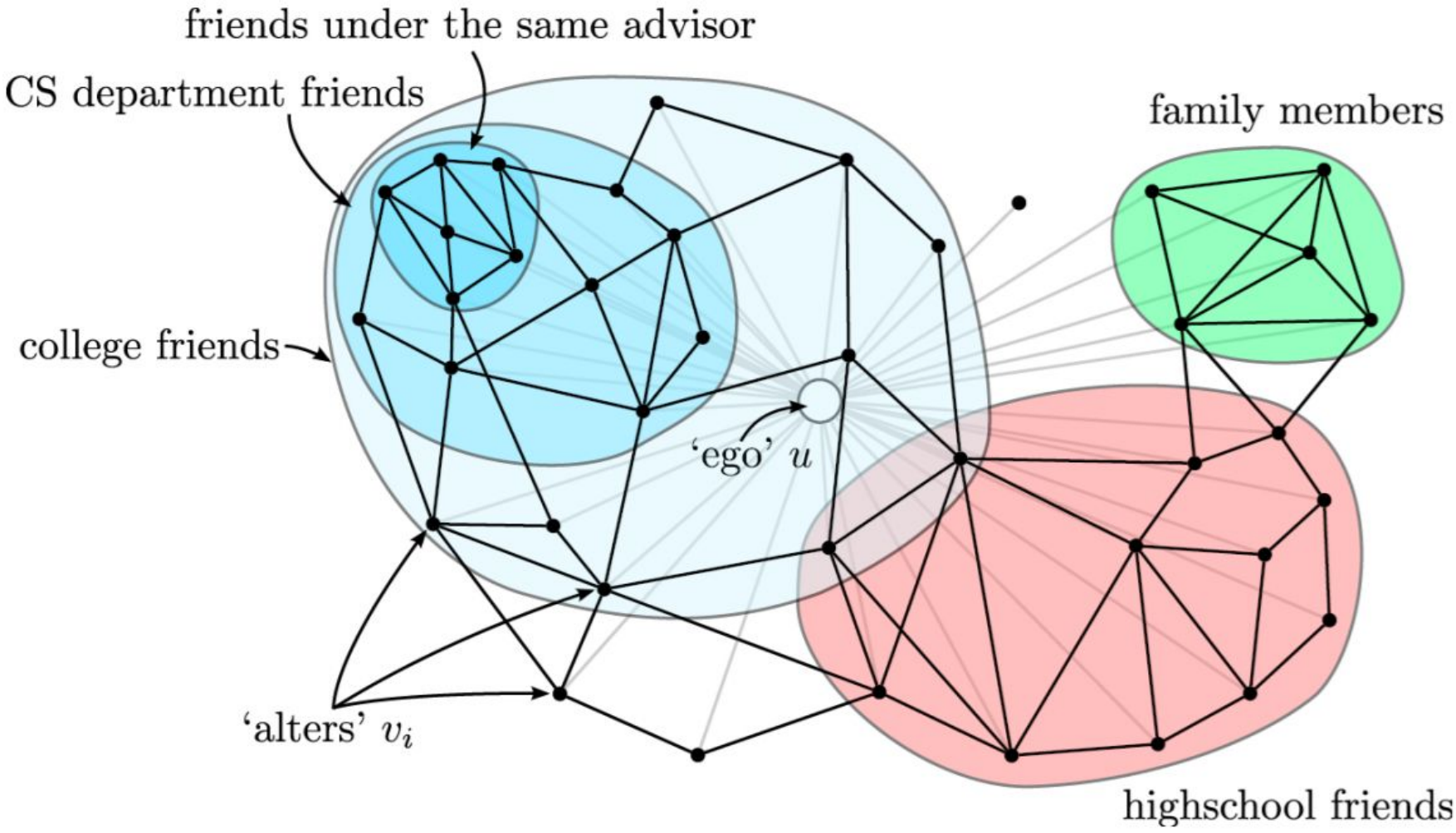
Results: 64 letters reached the target within 5.5 hops on average

Facebook case: Around 4 degrees of separation (https://arxiv.org/abs/1111.4570)

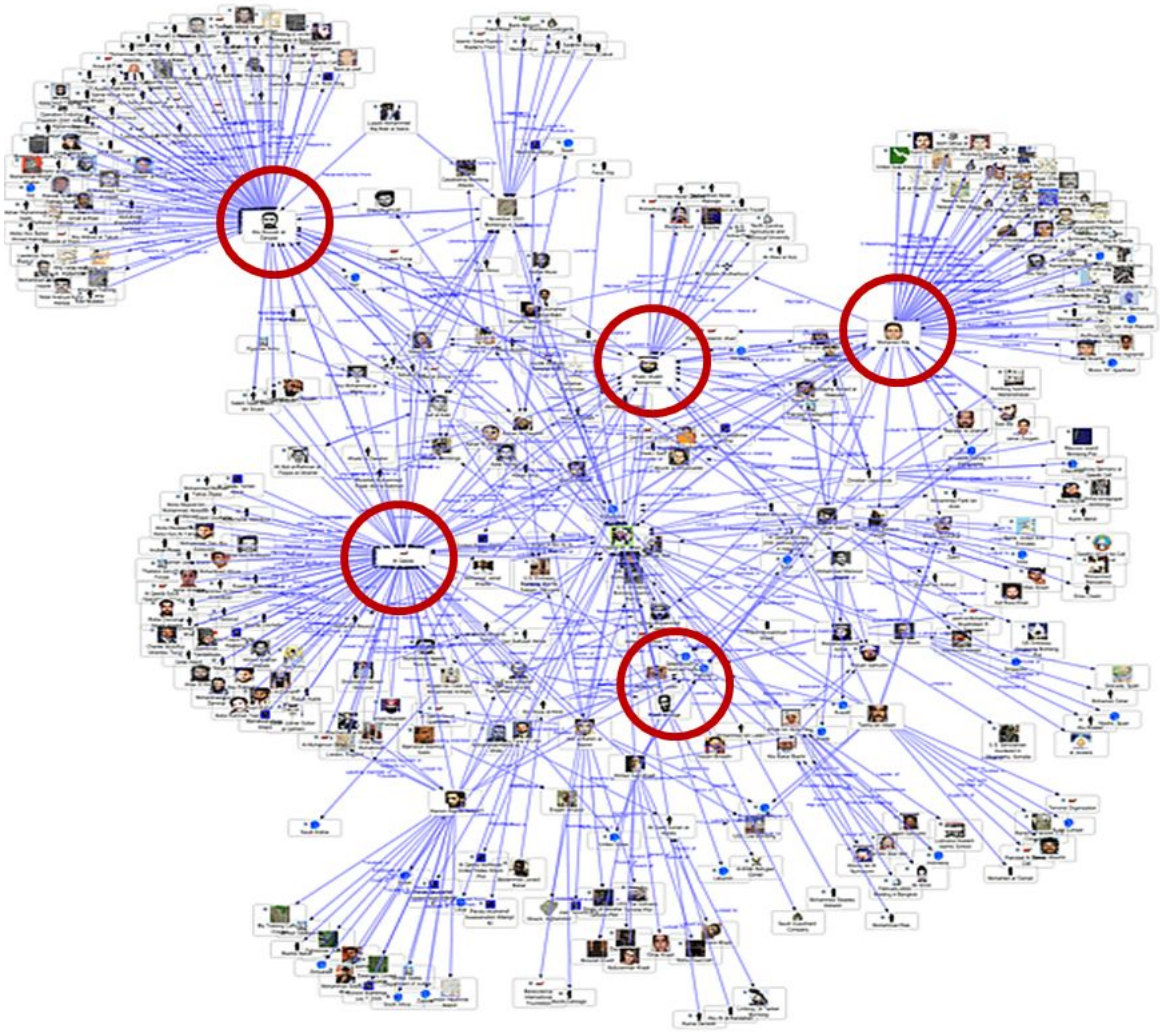
Community detection



Community detection



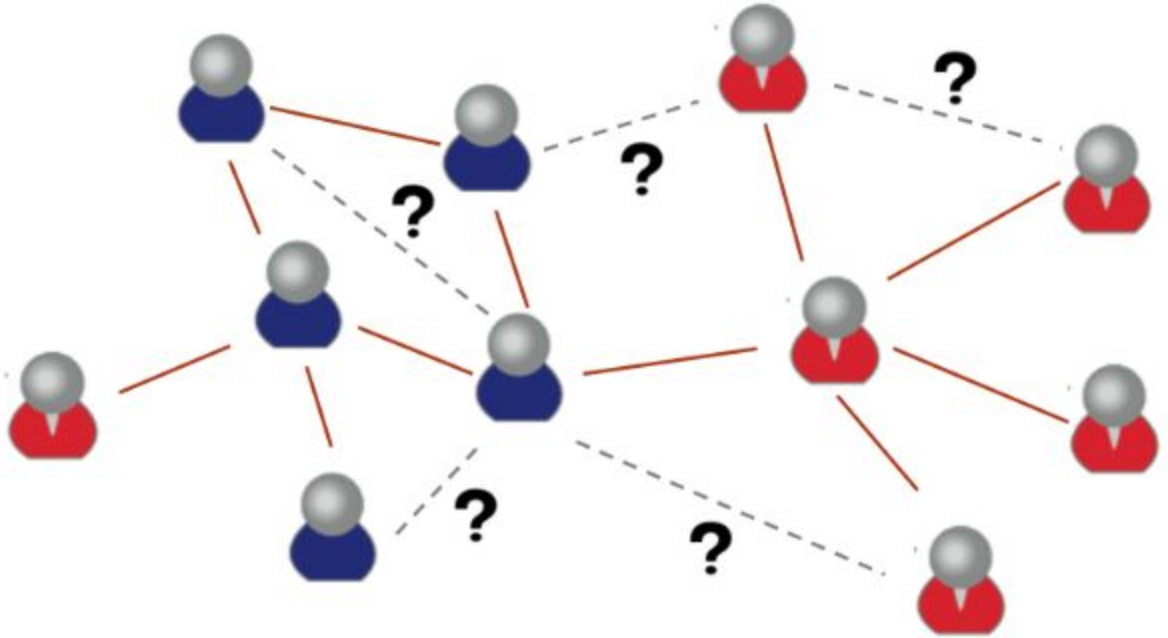
Influence Maximization



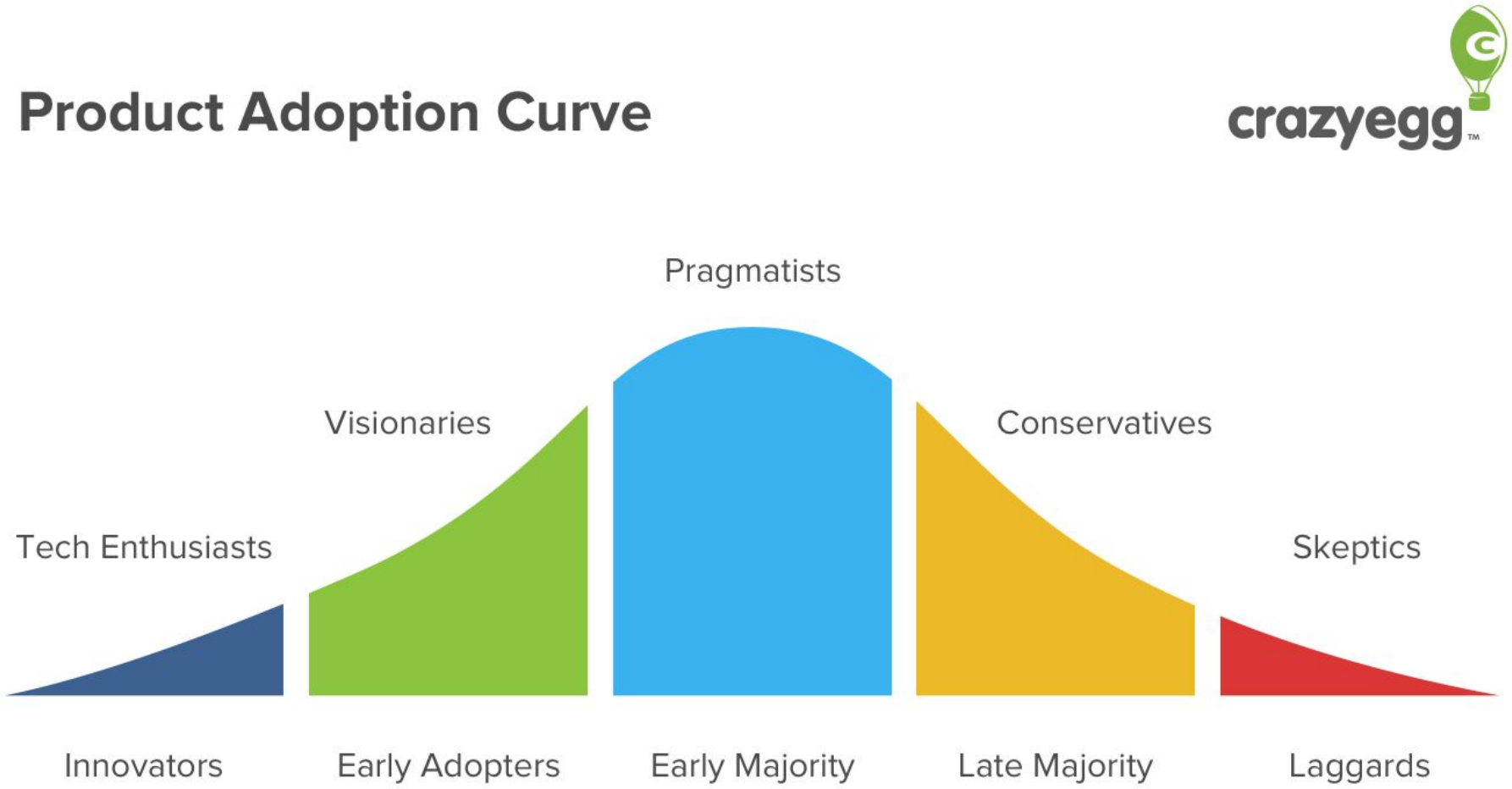
Find K individuals in the social network that maximize the influence

Link prediction

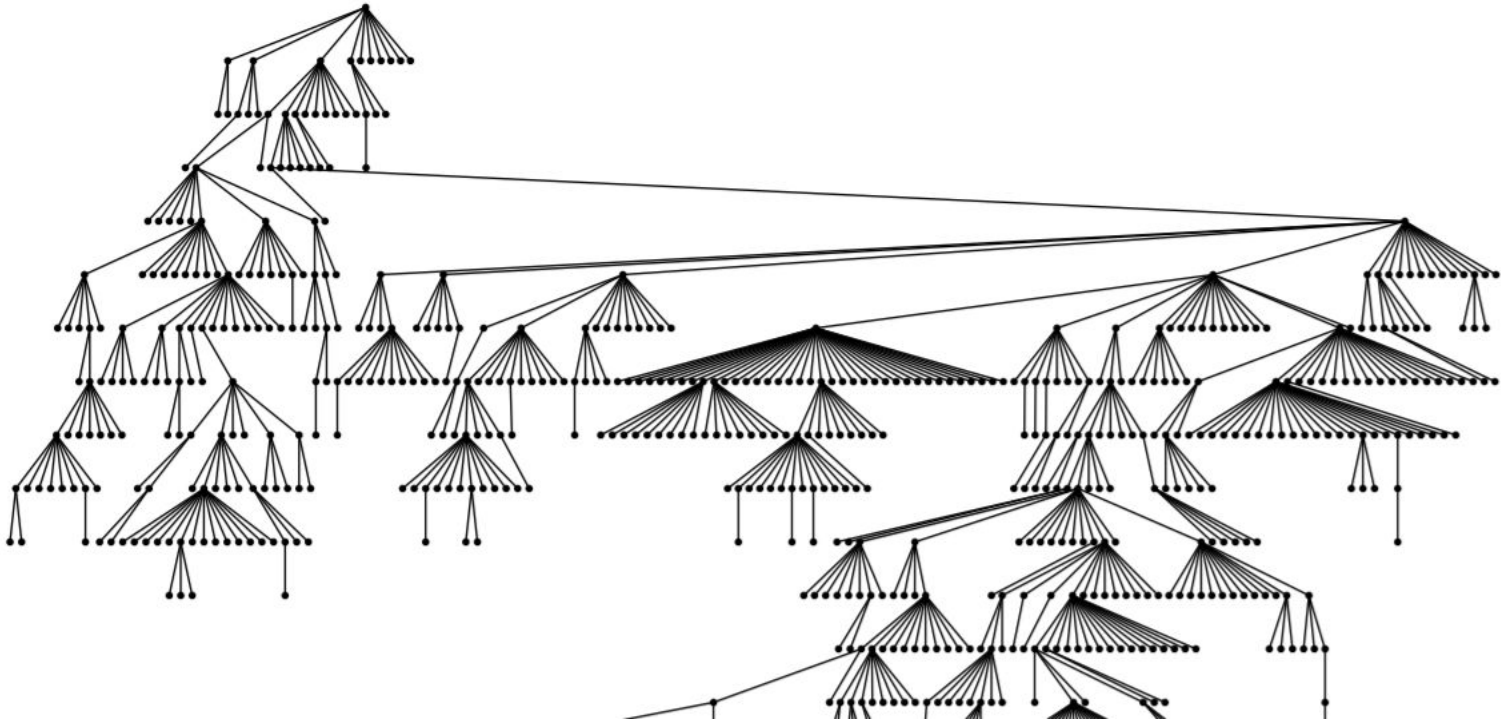
- “Suggested friends” feature



Product adoption



Product adoption



60% to 90% of LinkedIn users registered from friends invitation
(Anderson, Huttenlocher, Kleinberg, Leskovec, Tiwari, WWW'15)

Misinformation detection

Analyzing the content of the information and also the **source** and **pattern of spread**

Fake accounts detection

Detecting fake accounts using behavioral analysis

And more…

- Fraud financial activities
- Spread of diseases
- Employee and companies success
- …

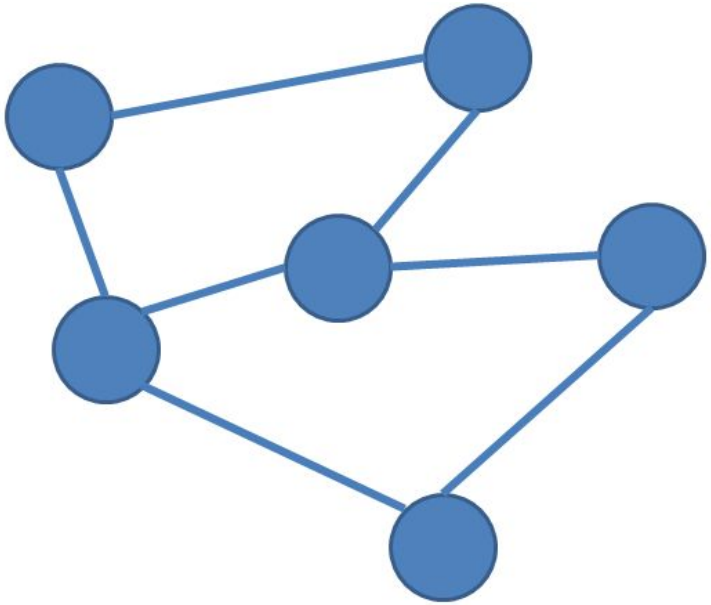
Summary

In this course we are going to focus on:

- Practical study of the data to find principles
- Mathematical models of the networks
  - Small-world model, structural balance,
- **Algorithms** (analyzing the network)
  - Communities detection, link prediction, influence maximization…
- **Applications**

Structure of the Network

Components of the Network



- **Vertices, Nodes** – objects/individuals **[V]**
- **Edges, Links** – interactions/relations **[E]**
- **Graph, Network** – the system **[G(V, E)]**

Modeling as Social Network

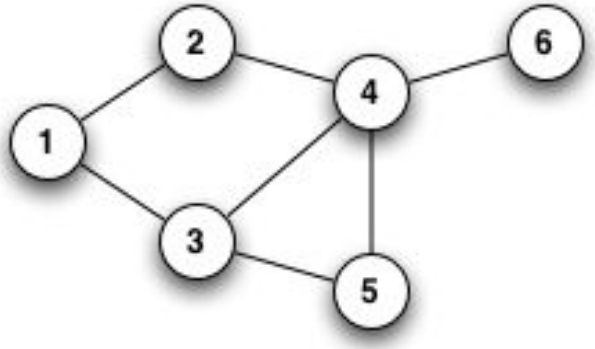
- Identify the domain:
  - Which problem you are trying to solve?
  - What are the nodes of the network?
  - What are the links of the network?

- .

Directed/Undirected Graphs

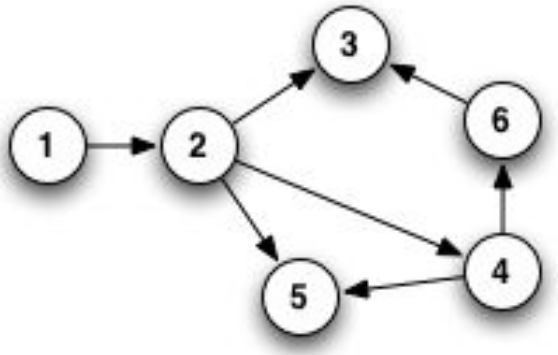
Undirected graph:

- Undirected, symmetrical edges
- Examples:
  - Friends (on Facebook)
  - Classmates



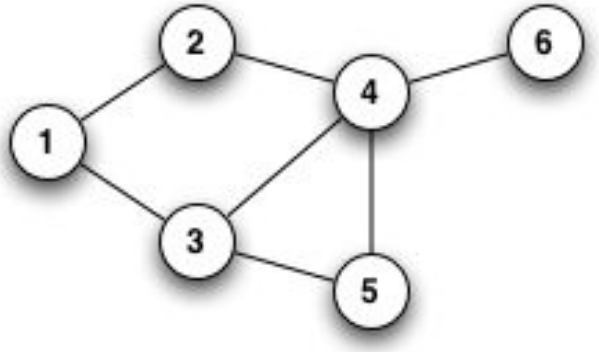
Directed graph:

- Directed edges
- Examples:
  - Followers (Instagram)
  - Phone calls



Node degree (Undirected)

Node degree (k_i) – number of edges adjacent to the node i



Example:

$k_5 = 2$, $k_3 = 3$

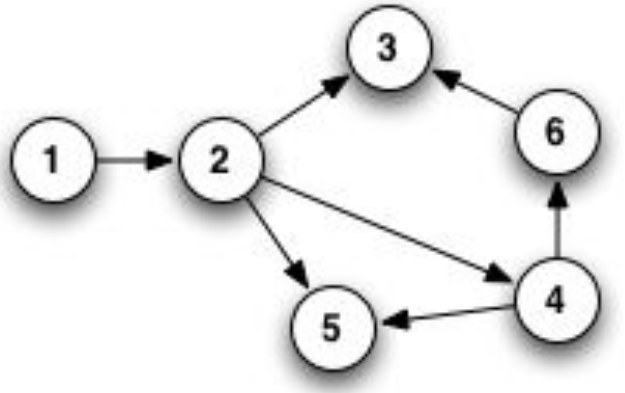
Average degree:

$$\langle k \rangle = 1/|V| * (k_1 + \ldots + k_{|V|}) = 2|E|/|V|$$

Node degree (Directed)

In-degree (k_i^{in}) – number of edges that goes from the node

Out-degree (k_i^{out}) – number of edges that goes from the node



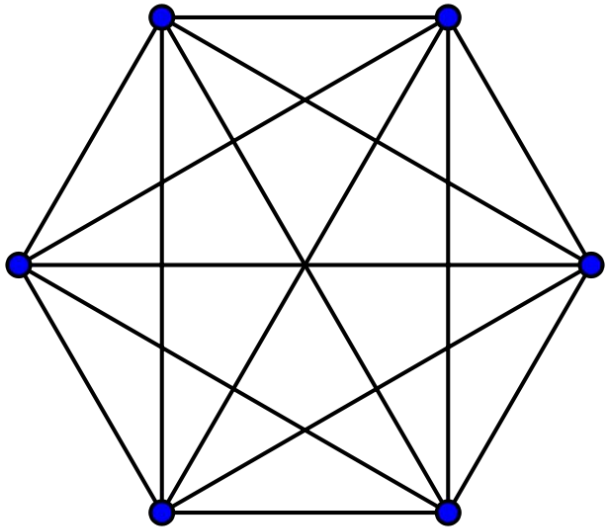
Total degree is a sum of in and out degrees.

Example:

$k_5^{in} = 2$, $k_5^{out} = 0$, $k_5 = 2+0=2$ $k_1^{in} = 0$, $k_1^{out} = 1$, $k_1 = 1$

Avg. degree: $\langle k \rangle = |E| / |V|$, $\langle k^{out} \rangle = \langle k^{in} \rangle$

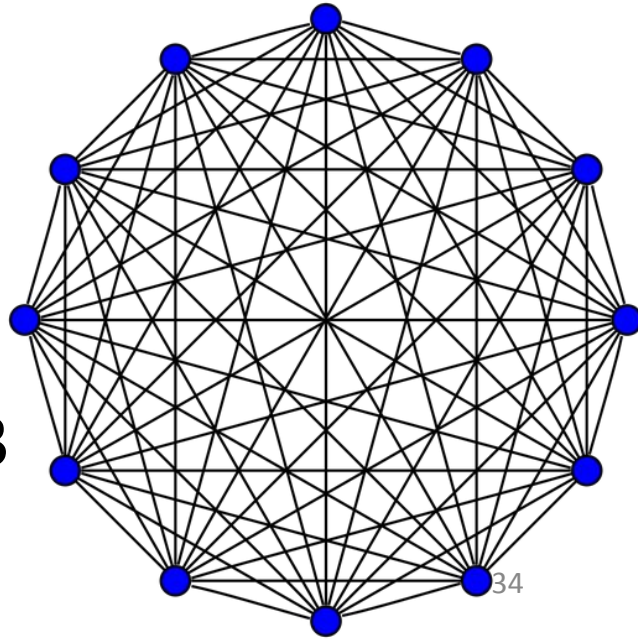
Complete Graph



The maximum number of edges in a graph of N nodes is

N*(N-1)/2

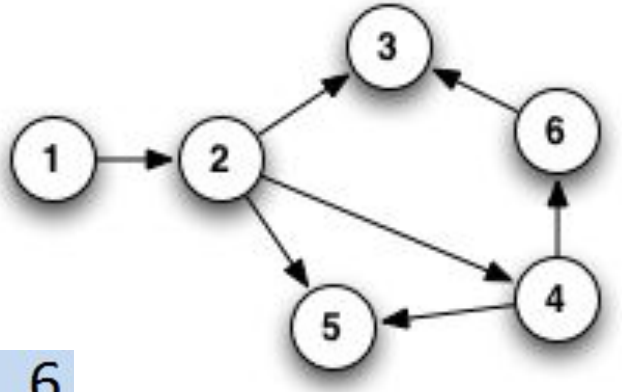
Undirected graph with maximum number of edges called **complete**



- clique is a complete subgraph
- triangle is a complete graph of size 3

Representing networks: Adjacency matrix

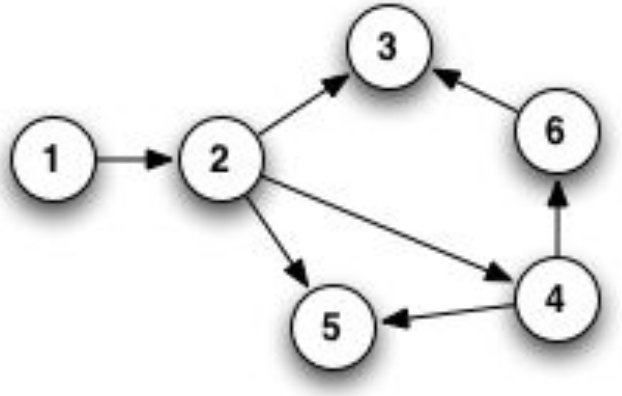
- A_{ij} = 1 , if there is an edge (i, j)
- A_{ij} = 0 , otherwise



	1	2	3	4	5	6
1	0	**1**	0	0	0	0
2	0	0	**1**	**1**	**1**	0
3	0	0	0	0	0	0
4	0	0	0	0	**1**	**1**
5	0	0	0	0	0	0
6	0	0	**1**	0	0	0

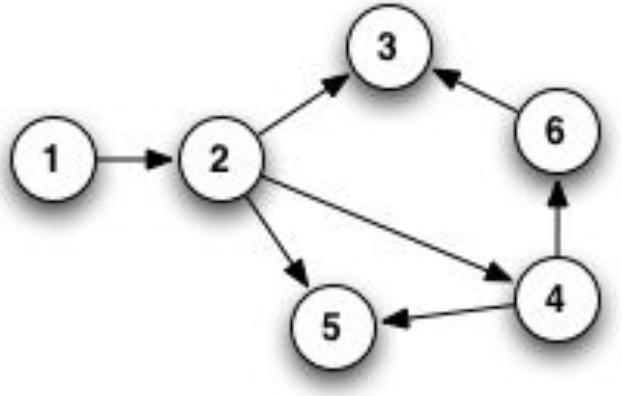
Representing networks: Edge list

- (1, 2)
- (2, 3)
- (2, 4)
- (2, 5)
- (4, 5)
- (4, 6)
- (6, 3)



Representing networks: Adjacency list

Easier for **large** and **sparse** graphs



- **1:** 2
- **2:** 3, 4, 5
- **3:**
- **4:** 5, 6
- **5:**
- **6:** 3

Social Networks are sparse

Most of the real world social networks are sparse

$$|E| << |E_{max}| \quad \text{or} \quad \langle k \rangle << |V| - 1$$

For example, in the LinkedIn social network:

$|V| \approx 7{,}000{,}000$ $\quad \langle k \rangle \approx 8.87$

(Source: Leskovec et al., Internet Mathematics, 2009)

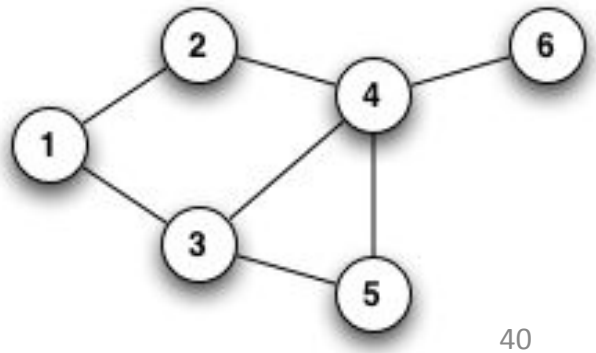
Edge attributes

- Weight (# messages, frequency of interaction)
- Ranking (most favorite actor, second favorite..)
- Type (friend, colleague, coauthor)
- Sign (positive/negative relationships)
- Properties depending on the other graph (number of common friends)

Connectivity of Undirected graphs

- **Connected graph** - any two nodes can be joined by a path (sequence of edges)
- **Disconnected graph** made out of 2 or more connected components

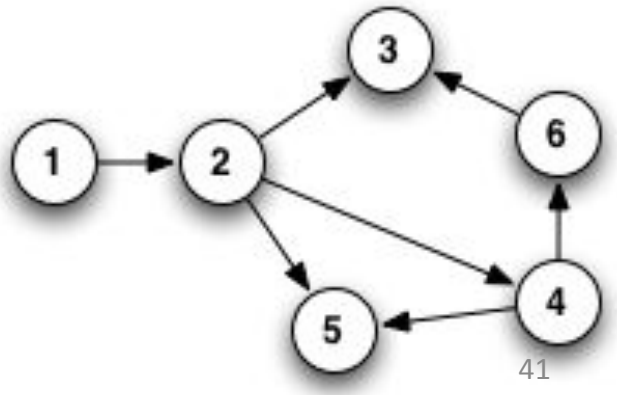
- **Bridge edge** – if we remove it, the graph becomes disconnected
- **Articulation node** - if we remove it, the graph becomes disconnected



Connectivity of Directed graphs

- **Strongly connected directed graph** – has a node from each node to each other node and vice-versa

- **Weakly connected directed graph** – connected if we ignore the edge directions



Quiz

For each of the examples, answer if the graph is directed/undirected and if edges are weighted or not

- Classmates –
- Facebook friends –
- Mobile phone calls –
- Twitter followers –
- Likes of Facebook –

Quiz

For each of the examples, answer if the graph is directed/undirected and if edges are weighted or not

- Classmates – undirected, weighted
- Facebook friends – undirected, non-weighted
- Mobile phone calls – directed, weighted
- Twitter followers – directed, non-weighted
- Likes of Facebook – directed, weighted

Network Properties

Key Network Properties

- Degree distribution P(k)
- Path length h
- Clustering coefficient C

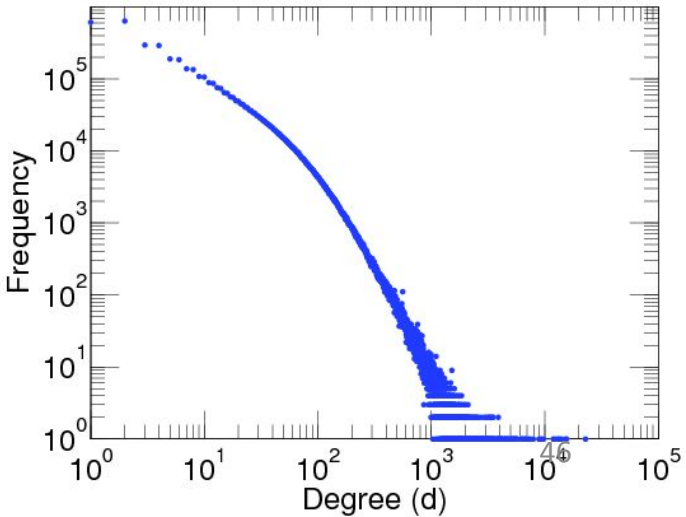
Degree distribution

- P(k) – probability that a randomly chosen node has a degree k

Given a graph with N nodes:

- $P(k) = N_k / N$ (N_k = # of nodes with degree k)

- Example of such distribution (LiveJournal)



Path length

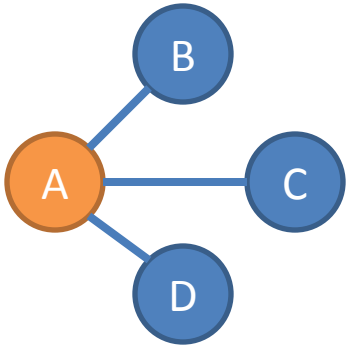
- **Path** - sequence of edges which connect a sequence of vertices which are all distinct

- **Distance** – the number of edges along the shortest path connecting two nodes

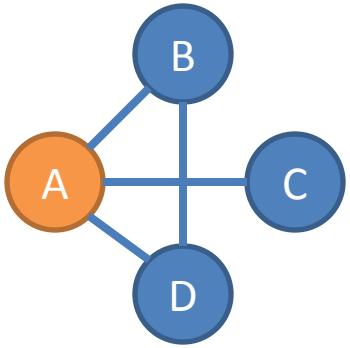
- **Diameter** – the maximal shortest path between two nodes in graph

Clustering coefficient

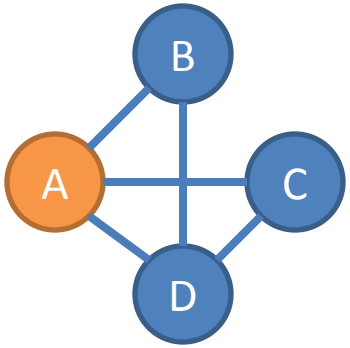
- **Clustering coefficient of a node** – fraction of the neighbors that are connected
- Node i, with degree k_i
- C_i = 2 * (# of edges between the neighbors)/ k_i *(k_i – 1)
- Intuitively: # of closed triangles / # of all triangles



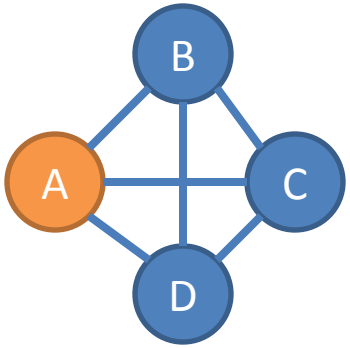
C_A = 0



C_A = 1/3



C_A = 2/3



C_A = 1

Clustering coefficient

- **Clustering coefficient of a node** – fraction of the neighbors that are connected

- Average clustering coefficient:

$$C = \frac{1}{N} \sum_{i}^{N} C_i$$

Thank you!

Questions?