Centrality measures in online social networks

Max Pellert, University of Konstanz

Social Media Data Analysis

Overview

1. Centrality and importance in social networks
2. The death of social networks
3. Coreness centrality and social resilience

The concept of centrality

Social network analysis can be used to measure the importance of a person as a function of the social structure of a community or organization.

In social networks, **centrality** measures this kind of structural importance of the node of a person. There are various centrality measures that stem from different kinds of structural importance.

Today you will learn about four centrality measures:

1. Degree centrality
2. Betweenness centrality
3. Closeness centrality
4. Coreness centrality

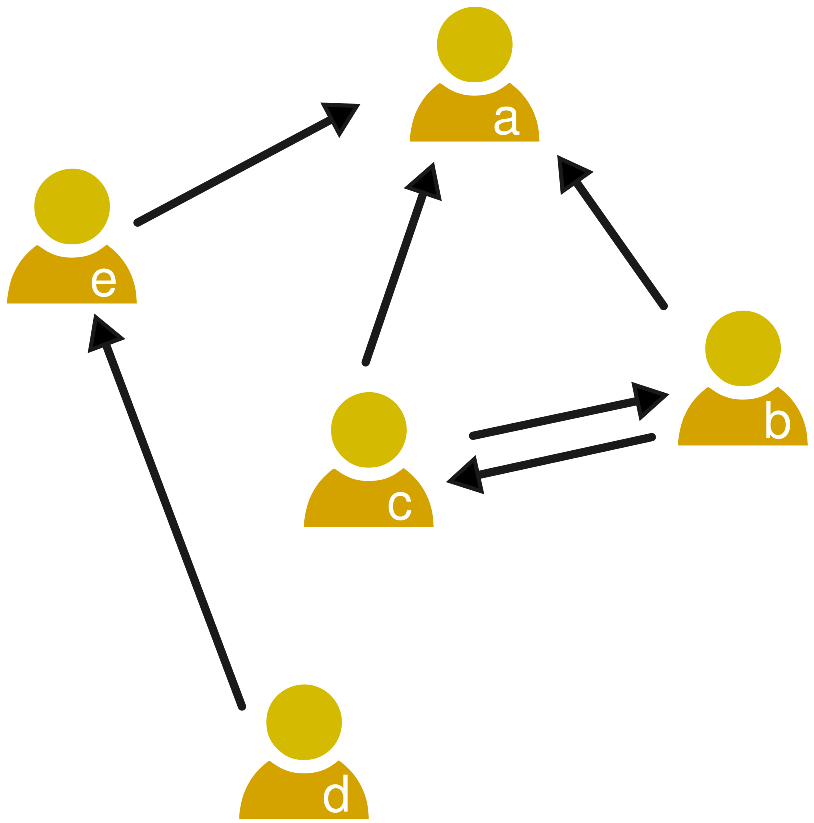
Degree centrality

A node's **degree centrality** measures the number of links connected to it.

In directed networks:

- **in-degree** $d_{in}(i)$ that is the number of edges ending in i, i.e. (j, i)
- **out-degree** $d_{out}(i)$ that is the number of edges leaving from i, i.e. (i, j)

$d_{in}(c) = 1$ and $d_{out}(c) = 2$



If importance on Twitter is the number of followers of an account, in-degree centrality is a way to measure it.

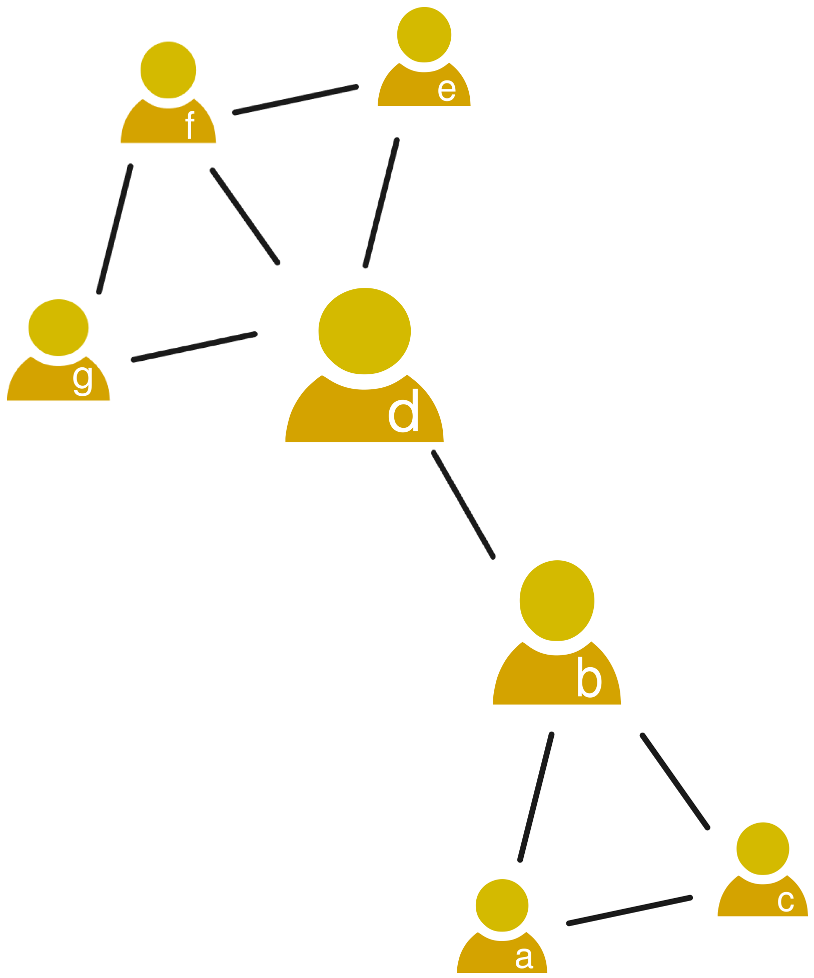
Betweenness Centrality

Sometimes the importance of a person is quantified as the number of shortest paths between two other people that pass through this person. In this case, **betweenness** centrality measures importance:

$$C_B(i) = \sum_{s \neq i, t \neq i} n_i(s,t)$$

Where $n_i(s,t)$ is the number of shortest paths from s to t that pass through i.

Example: $C_B(b) = 16$



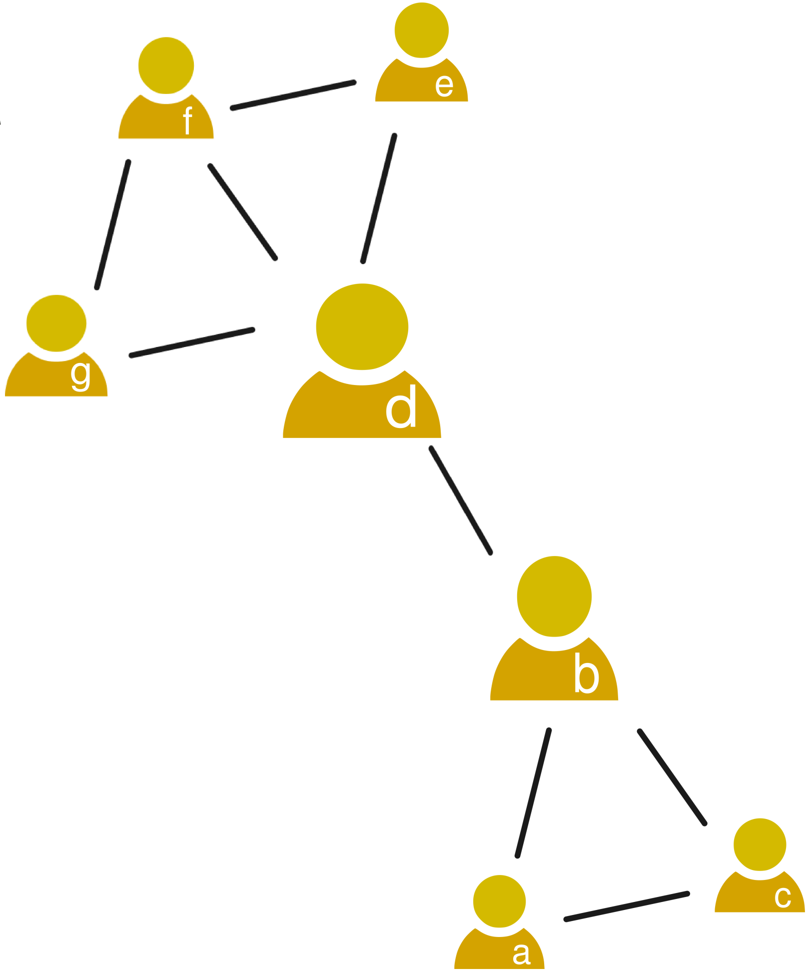
Closeness Centrality

Sometimes the most important people in a group are the ones that can reach everyone with the least effort. In these cases, **closeness** centrality measures importance as:

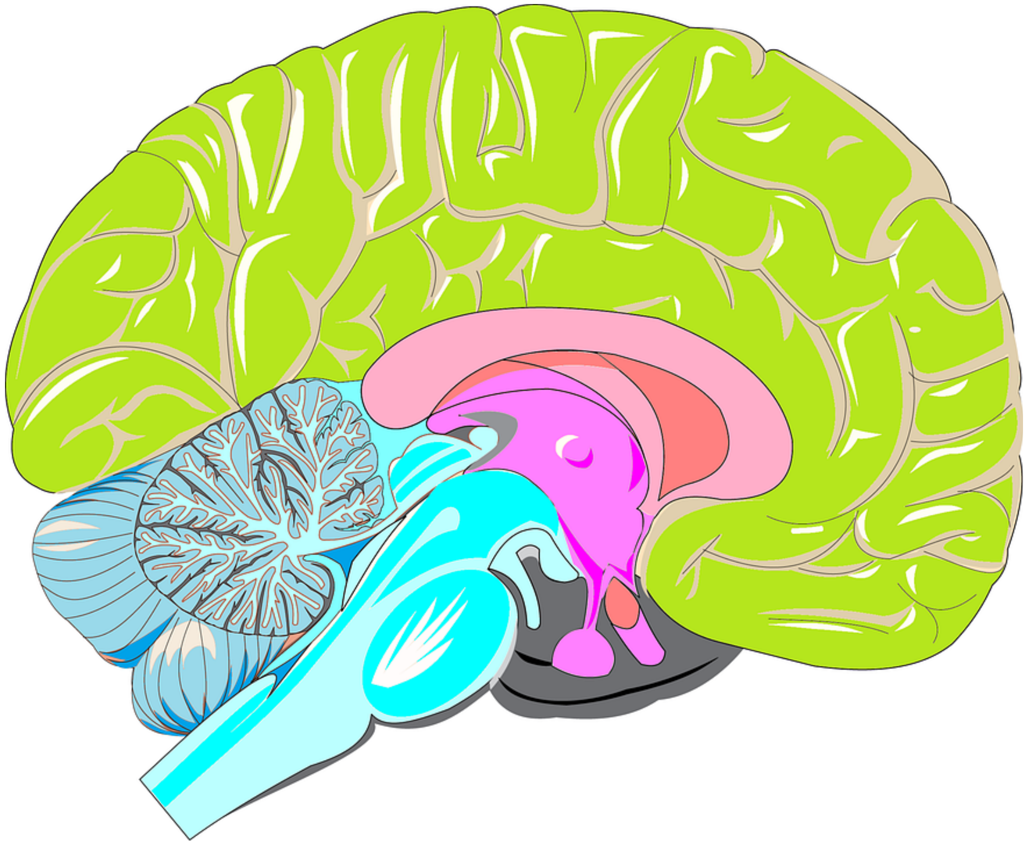
$$C_C(i) = \frac{n-1}{\sum_{j \neq i} dist(i,j)}$$

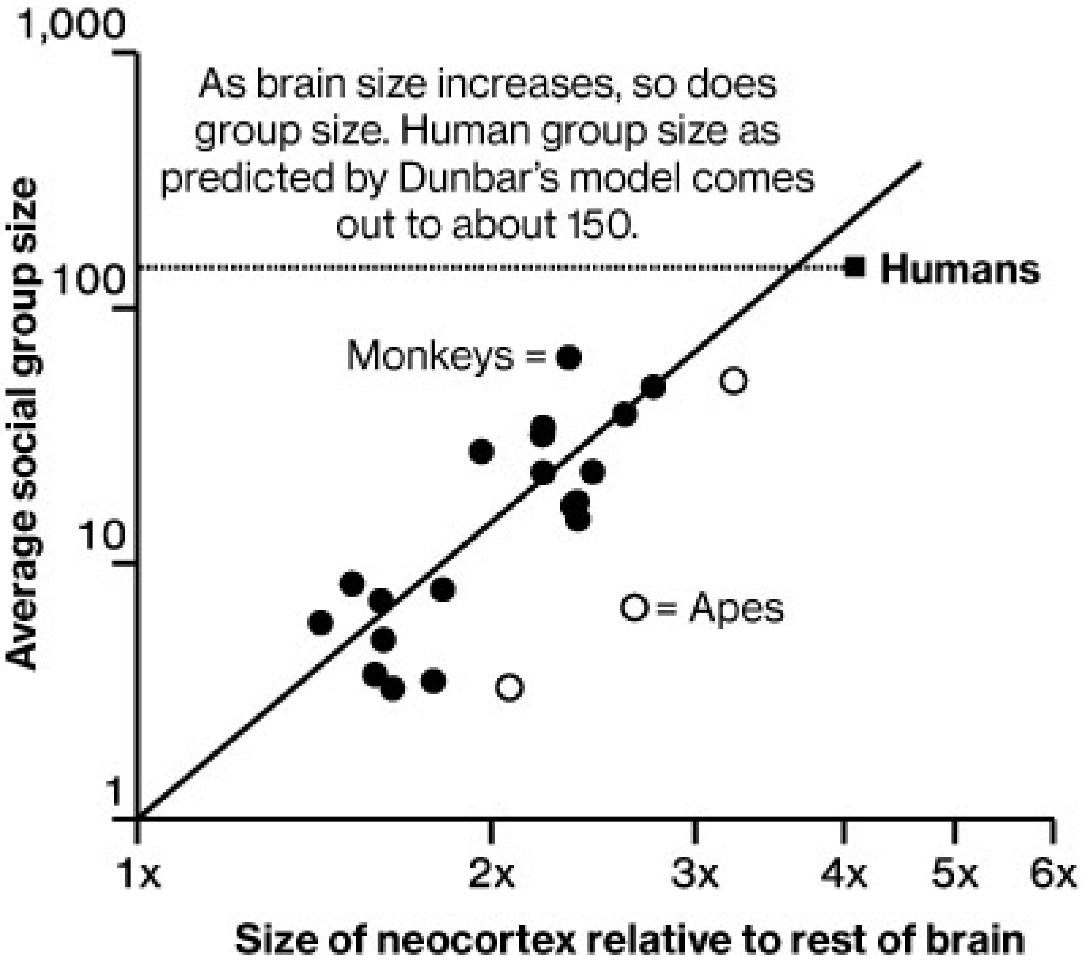
Where $dist(i,j)$ is the distance from i to j and n is the number of nodes in the network.

Example: $C_C(d) = 0.75$

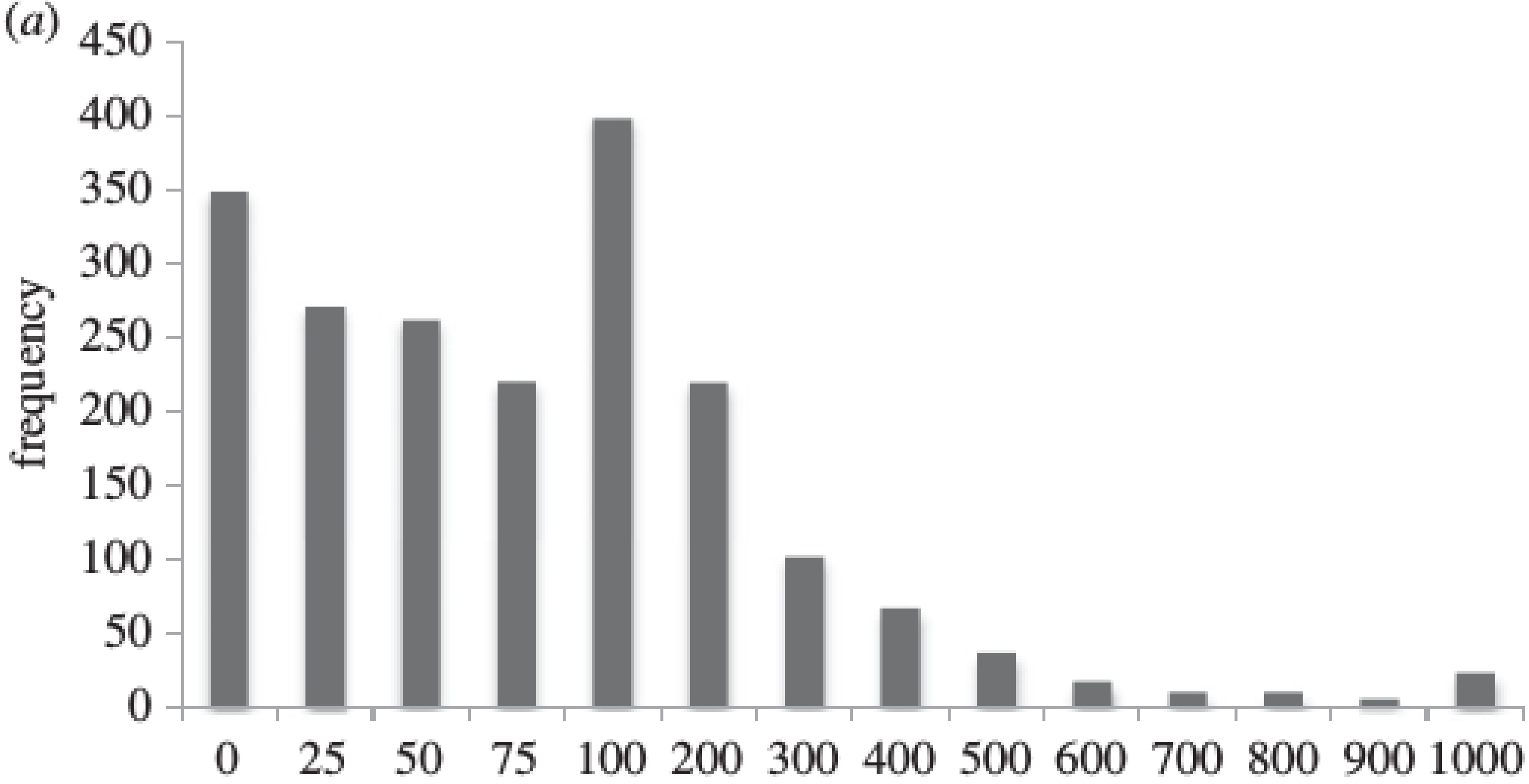


Neocortex Size as a limit to degree centrality

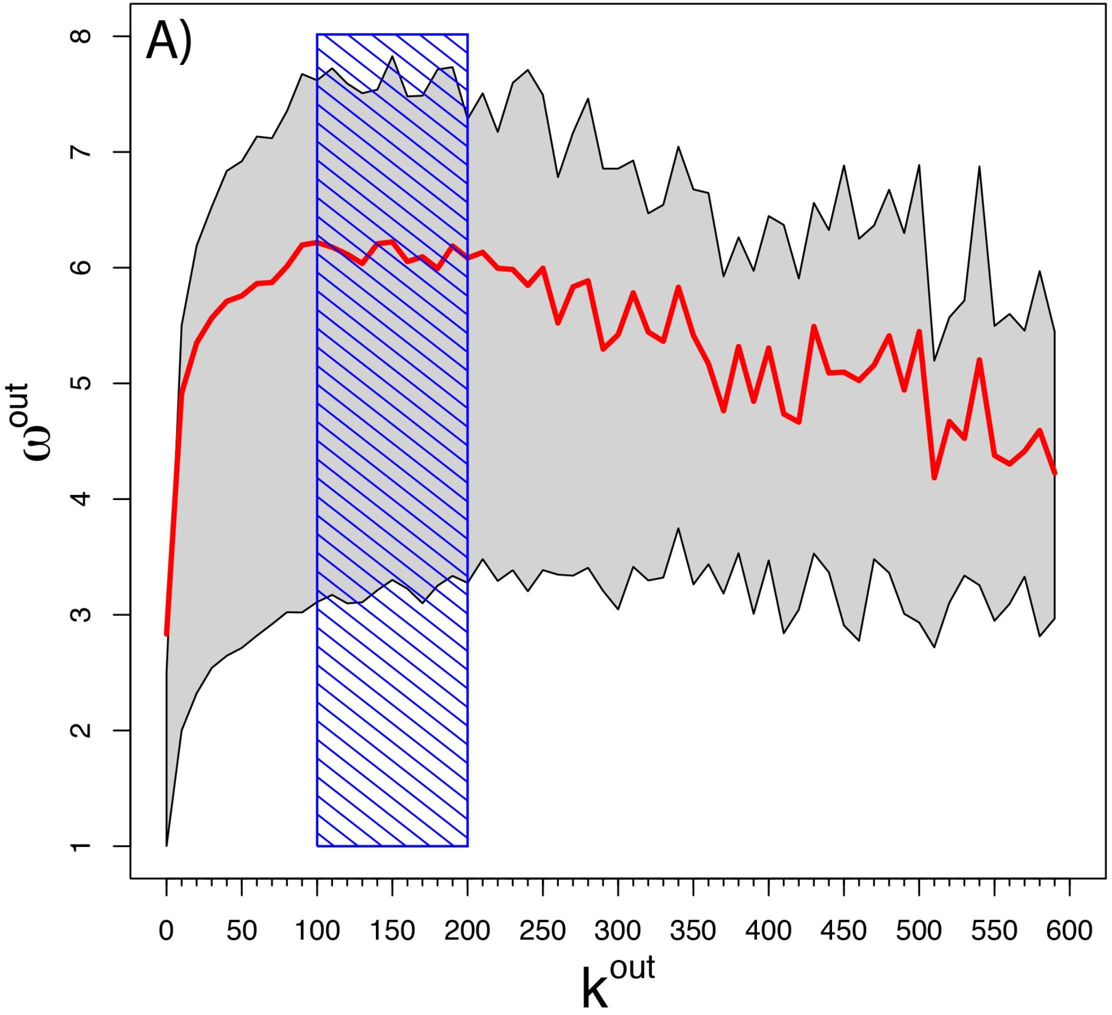




Dunbar's number in Online Social Networks



Dunbar's number on Twitter

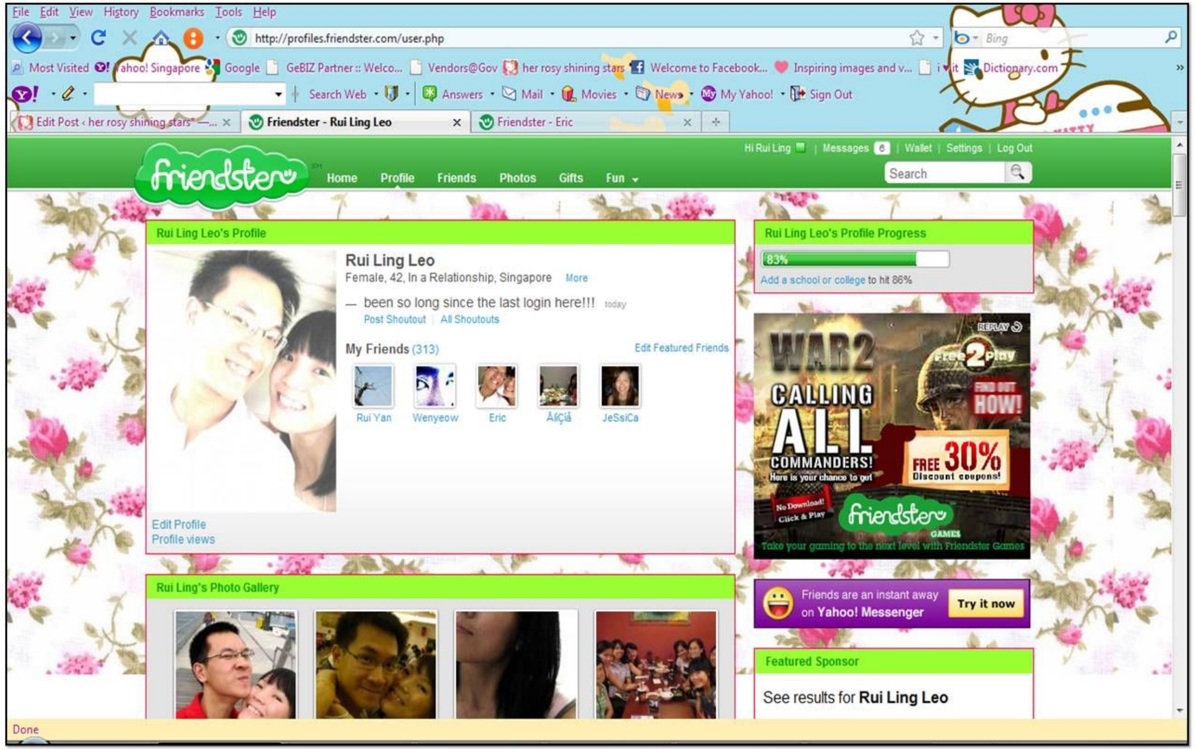


The death of social networks

1. Centrality and importance in social networks
2. The death of social networks
3. Coreness centrality and social resilience

Why do online social networks collapse?

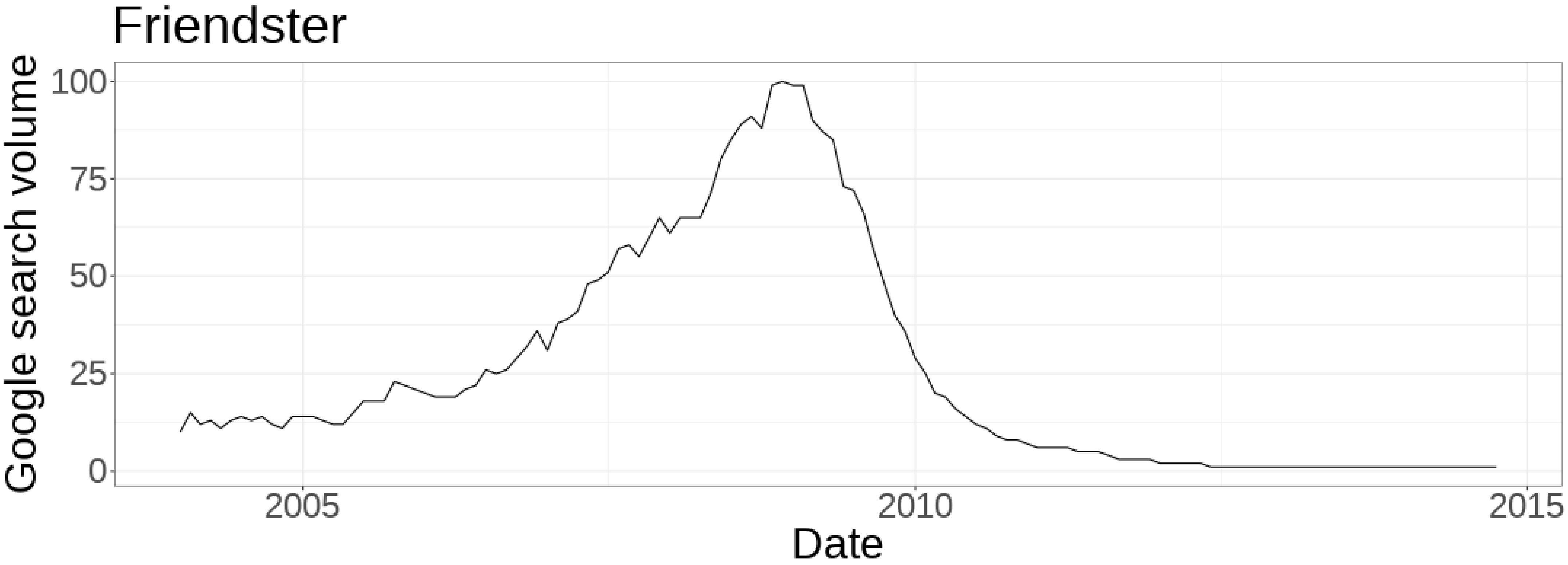
> **Social Resilience:** The ability of a community to withstand external stresses, disturbances, and environmental changes



Friendster Was the Best Social Media Site, Ever. Yudhistira Agato, Vice (2017)

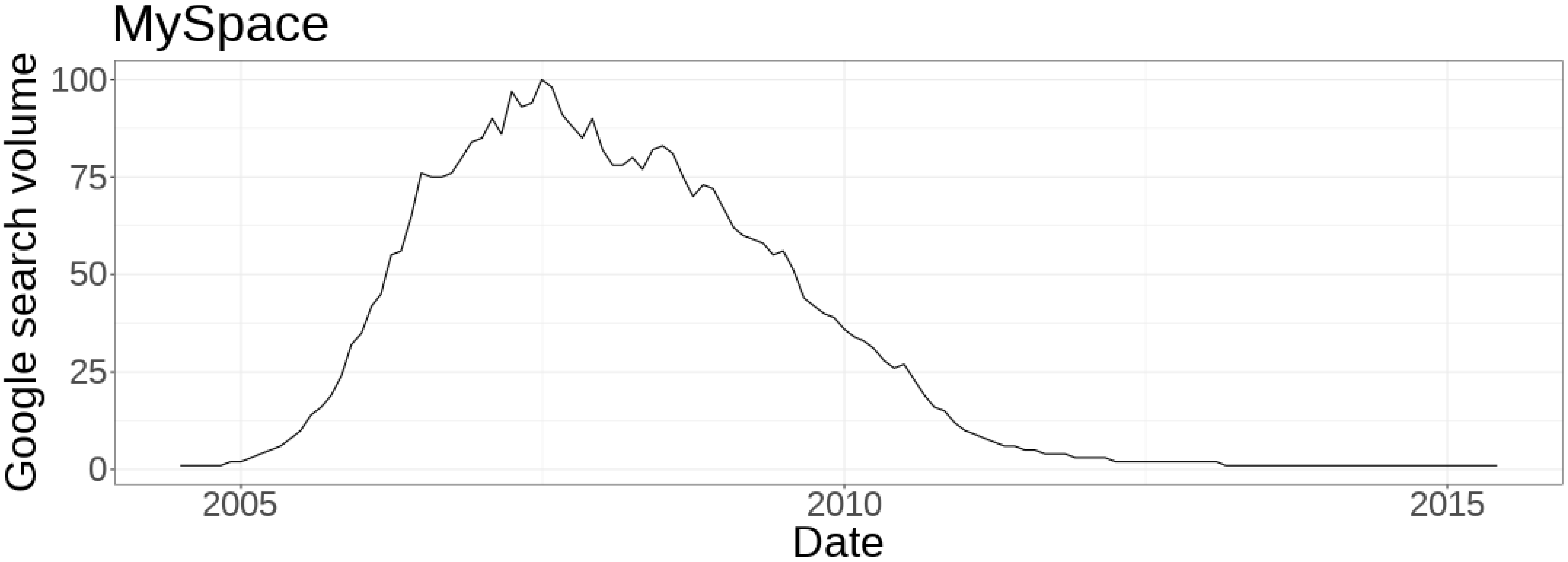
The rise and fall of Friendster

Friendster went from 80 Million active users to disappear completely.



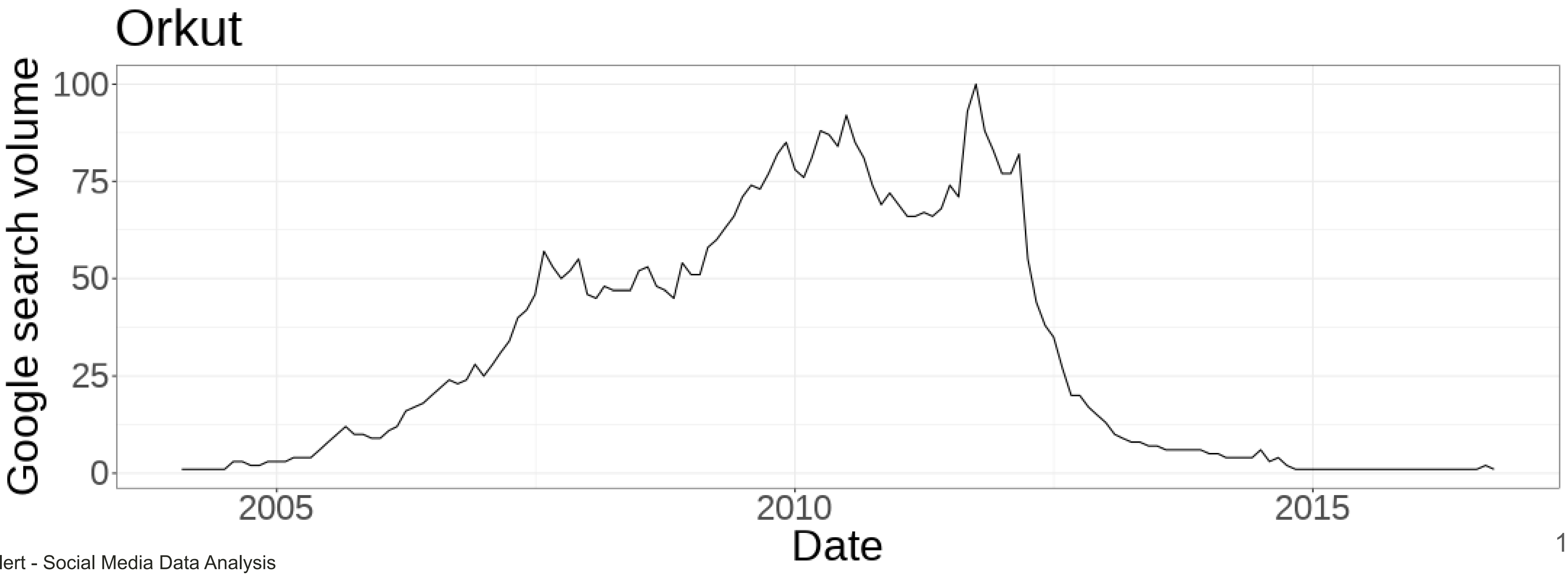
The rise and fall of MySpace

MySpace went from being valued more than 12 Billion USD in 2008 to be bought by Justin Timberlake for 35 Million USD.



The rise and fall of Orkut

Orkut was the first attempt of Google to launch an online social network. It was very popular in some countries but lost users to Facebook and it was eventually taken offline.



Coreness centrality and social resilience

1. Centrality and importance in social networks

2. The death of social networks

3. Coreness centrality and social resilience

Modelling social resilience online

Social resilience can be modelled as a process of how users stay active or inactive in a social networking site.

If we consider social network users as rational, they will respond to incentives to stay active or to abandon social networks depending on **benefits and costs**.

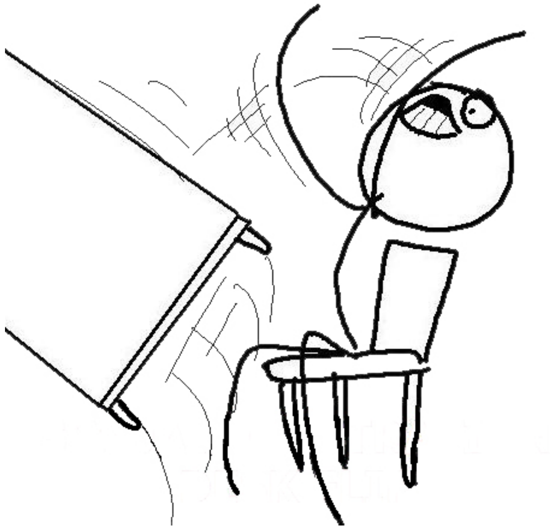
Benefits can be quantified through the content users receive from their friends (shares, comments) and through the attention and support given by their friends (likes, votes).

A simple way to model monotonic benefits is proportionally to the active friends of a user k_u:

$$benefit_u = b * k_u$$

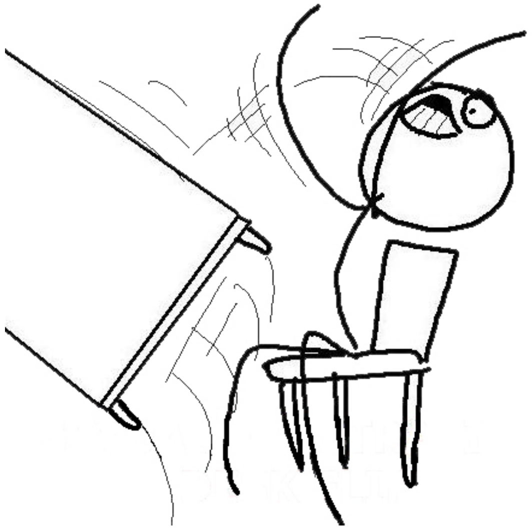
Costs of using an online social network

Costs: Using social network is not only benefits, there are also costs associated with being active.



Costs of using an online social network

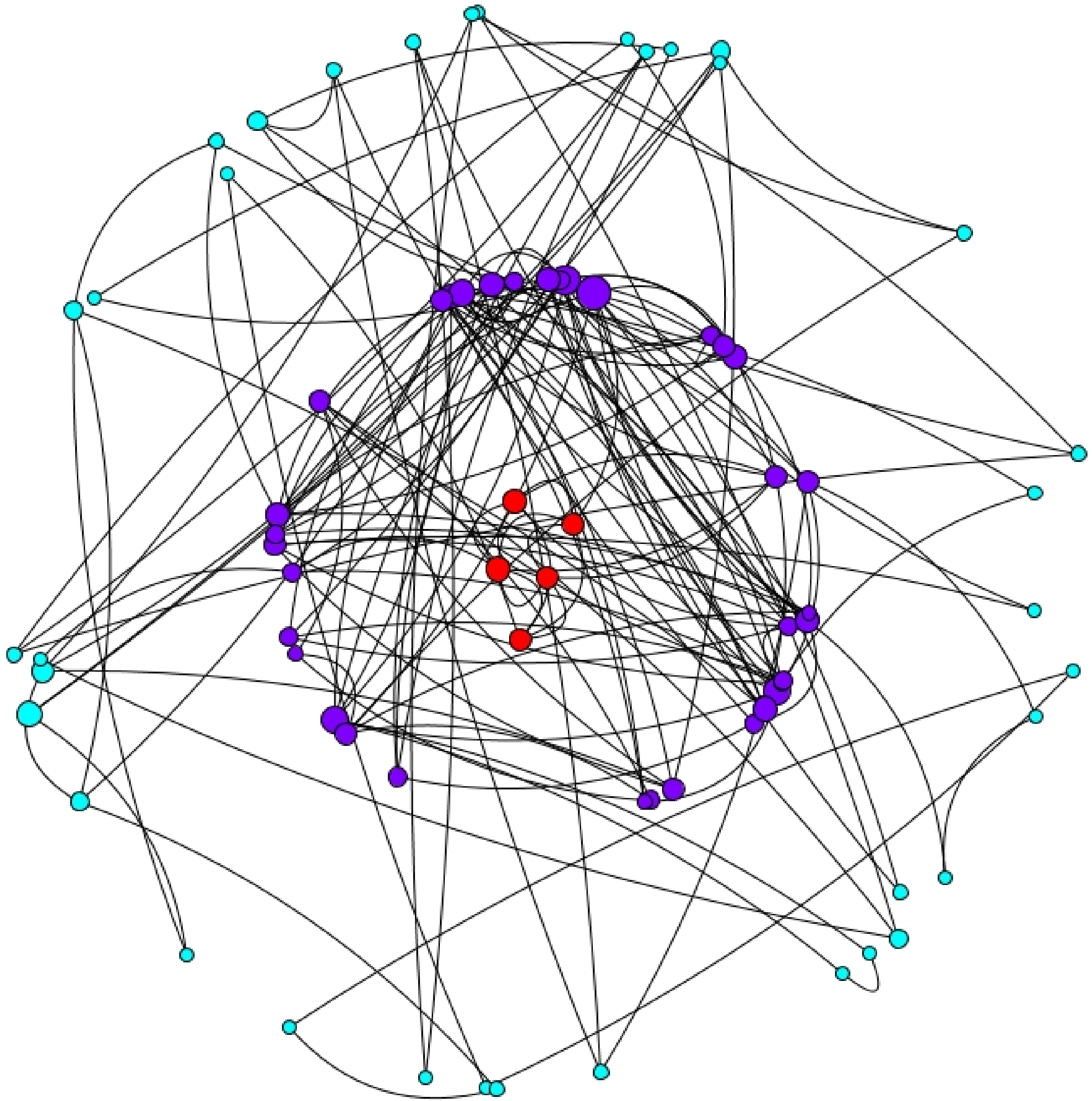
Costs: Using social network is not only benefits, there are also costs associated with being active.
For example:

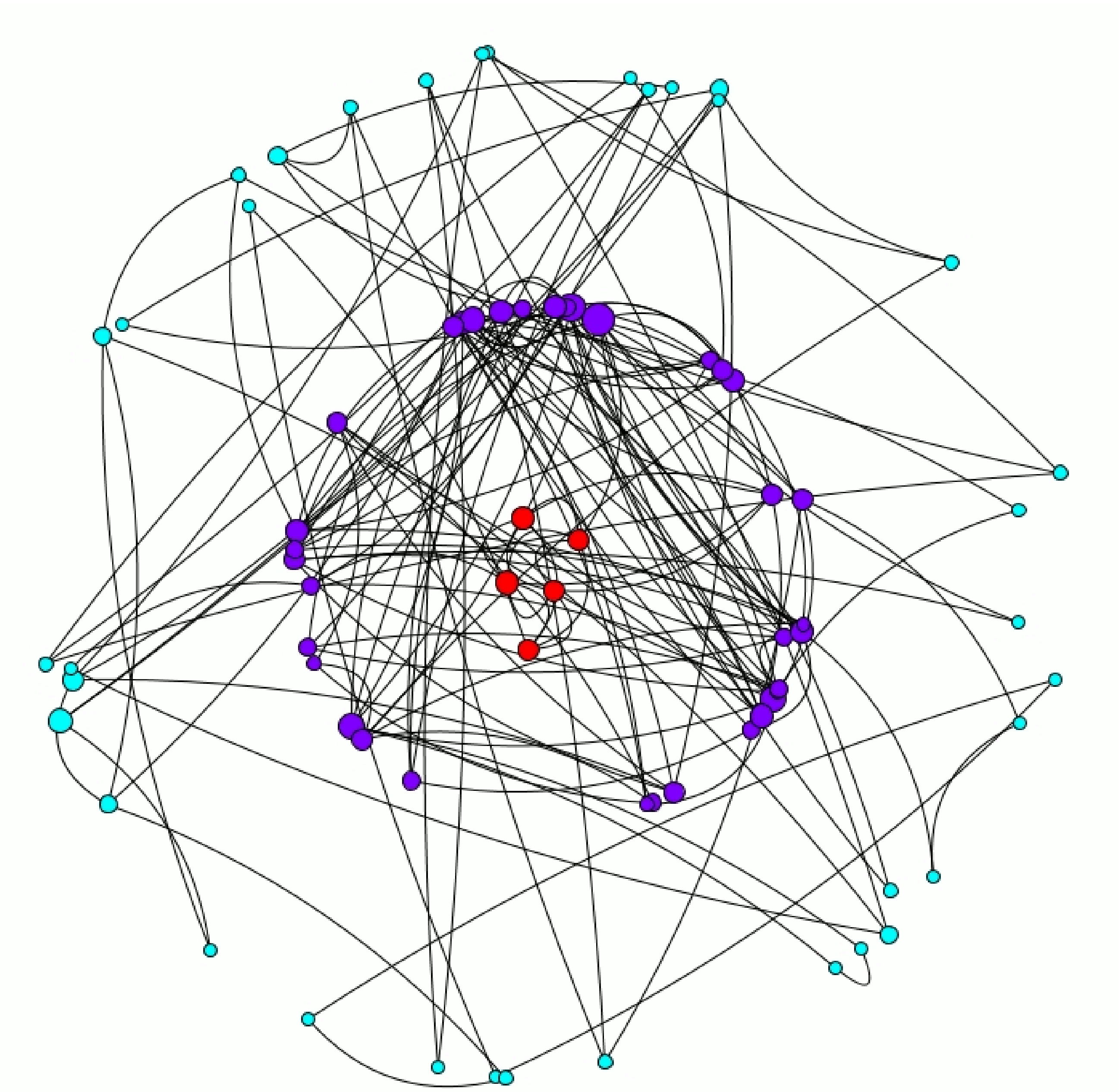


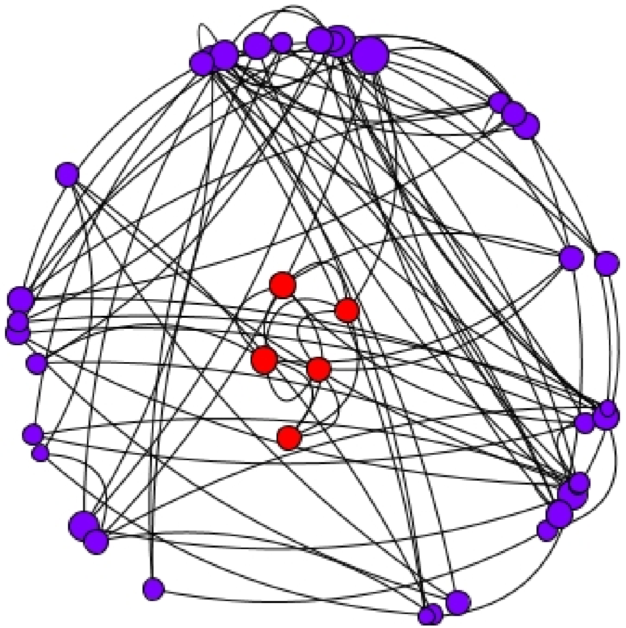
- Time spent to learn to use the interface of the platform
- Risks of disclosing personal information
- Opportunity costs: you could be doing something else
- Economic costs, for example membership fees.

A common assumption about costs is that they are relatively similar for all users, thus modeling them as a constant c. Then, a user will become inactive when the total benefits are below the cost:

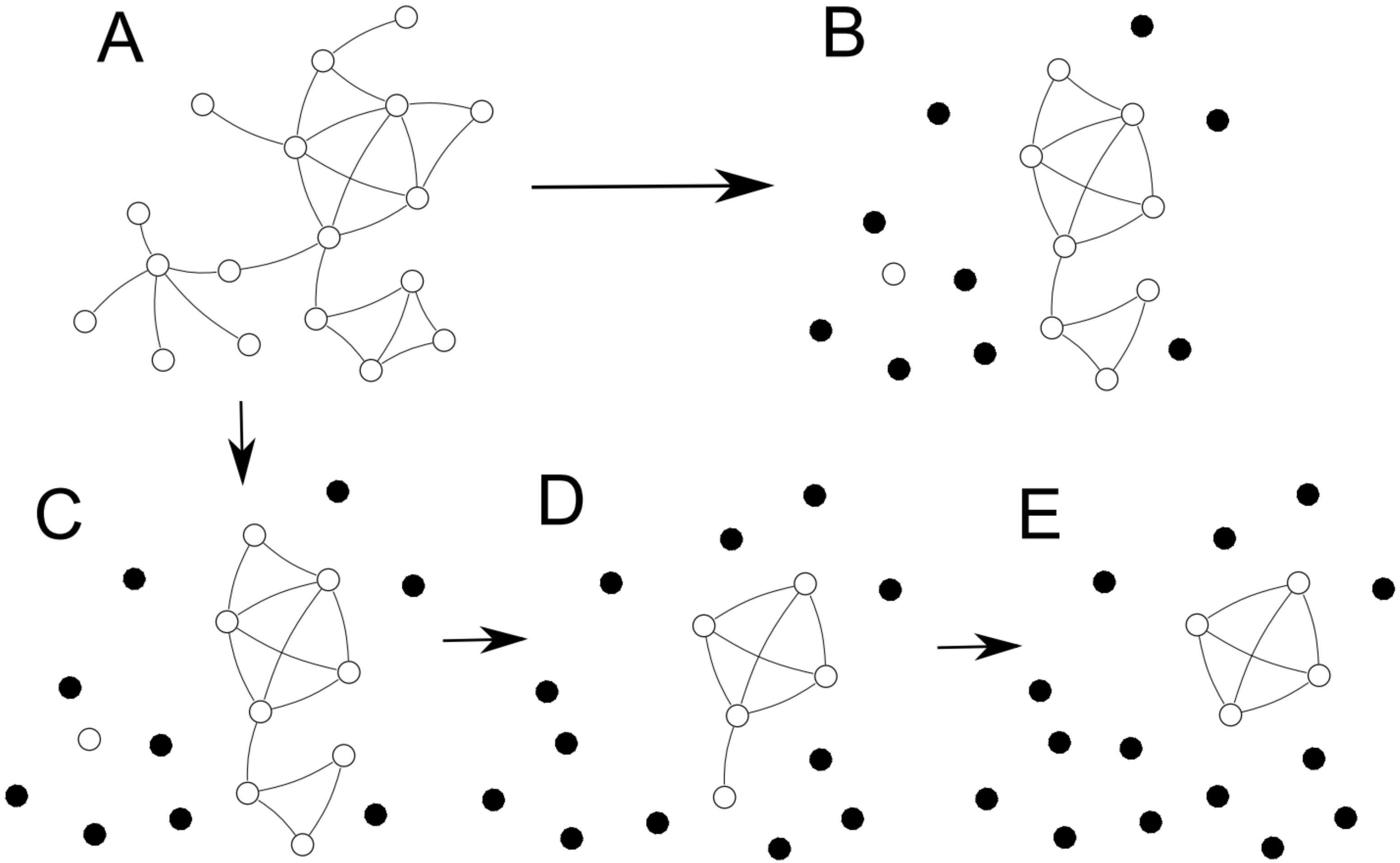
$$b * k_u < c$$







Iterative removal by degree



The k-core decomposition

The graph remaining after the cascade above is what is called a k-core

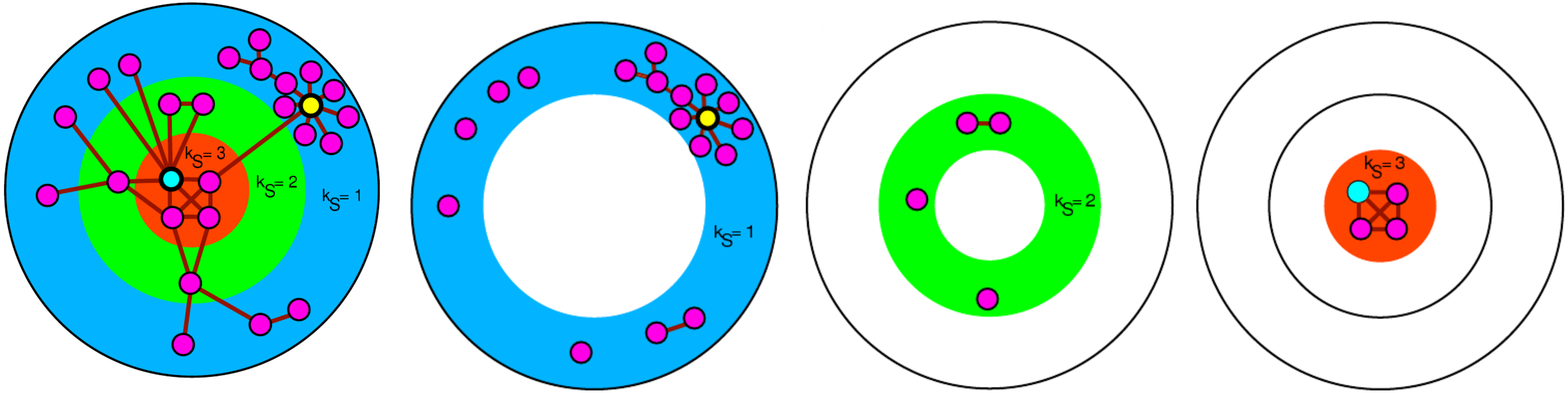
> **k-core**: A k-core of a graph G is a maximal connected subgraph of G in which all vertices have degree at least k.

For any network, you can calculate its k-core decomposition as follows:

- Start with $k_s = 1$
- Remove all nodes with degree less than or equal to k_s and their links
- Repeat until all nodes have degree larger than k_s
- Increase k_s by one and repeat until no nodes are left

The nodes and the edges removed for certain of $k_s = k$ is called the **k-shell**.
A **k-core** is the set of all k-shells with $k_s \geq k$.

Coreness centrality

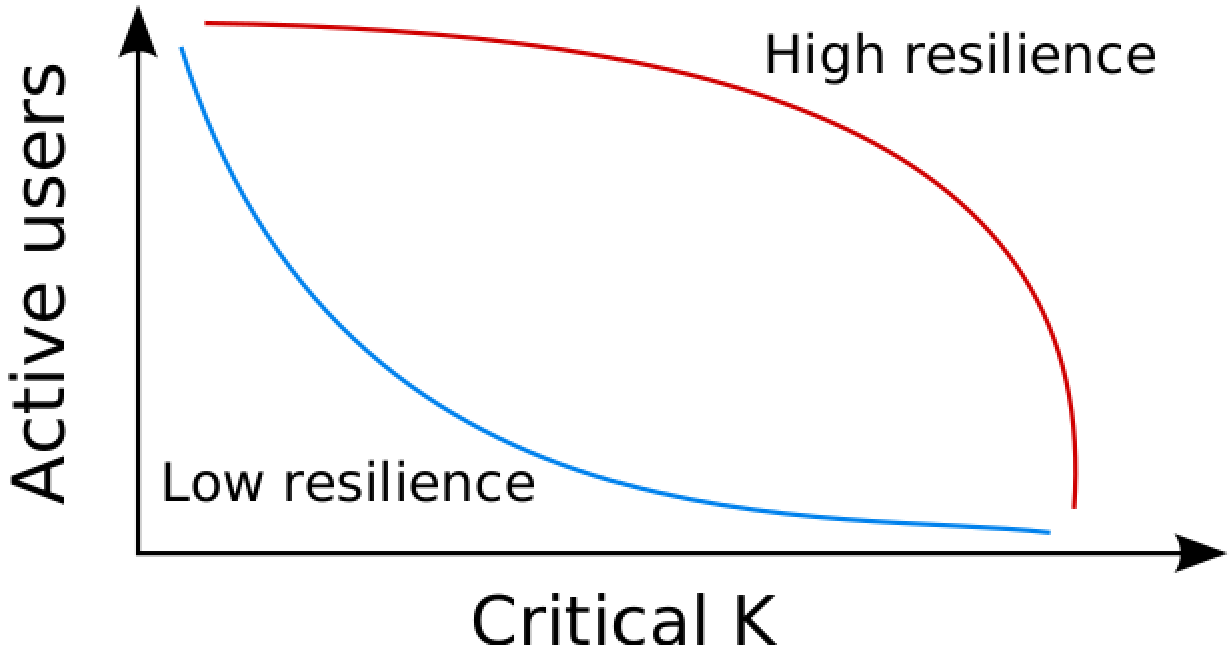


The k-shell number of a node is also called **coreness centrality**.

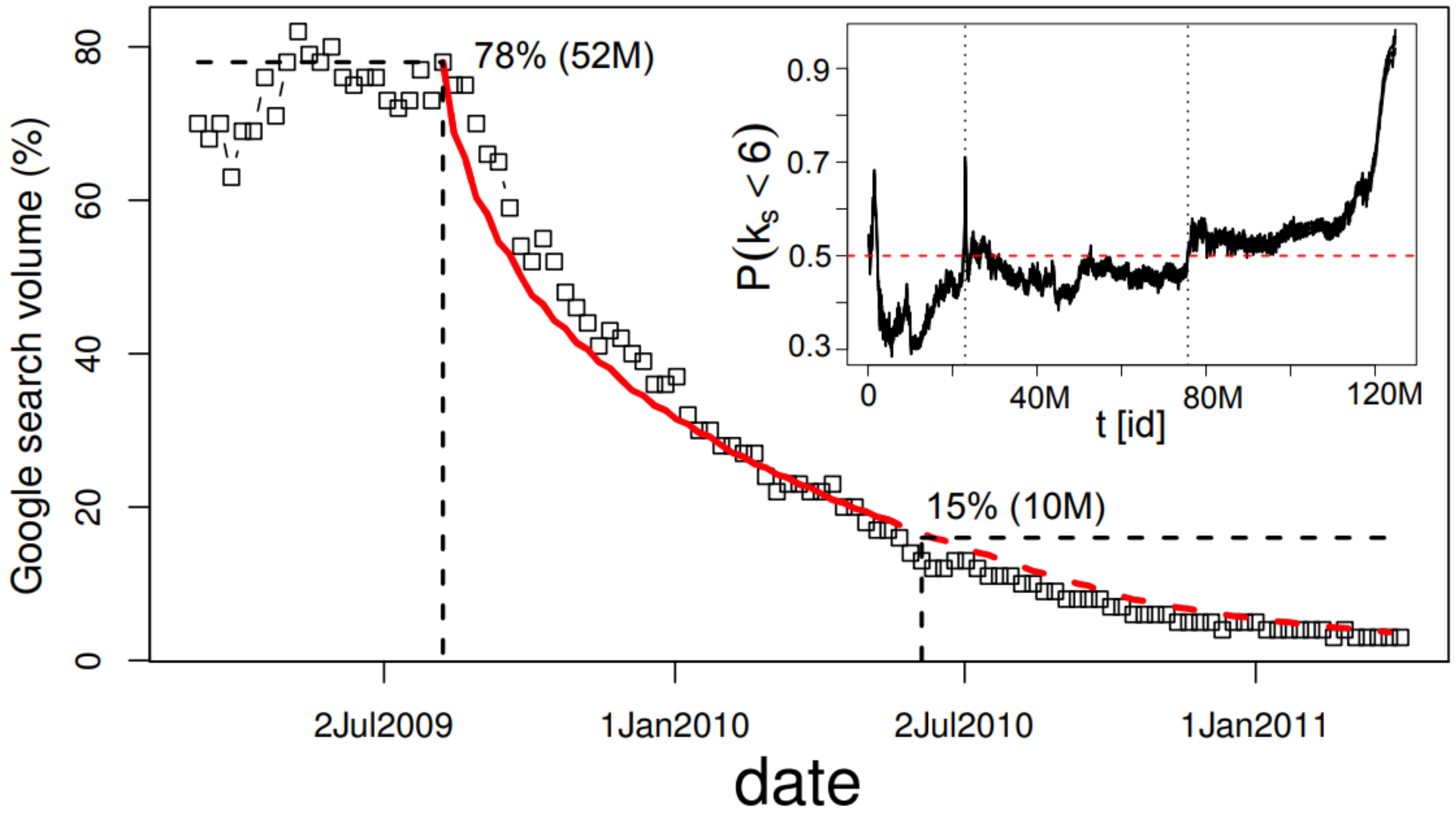
Figure from Kitsak et al. Identification of influential spreaders in complex networks. Nature Physics (2010)

Coreness and social resilience

The cost to benefit ratio c/b defines a critical value of the degree K, below which users with degree $k_u < K$ will leave the social network. The remaining active social network is the k-core corresponding to K. The **cumulative density function** of coreness values in the network serves as a **resilience function**:

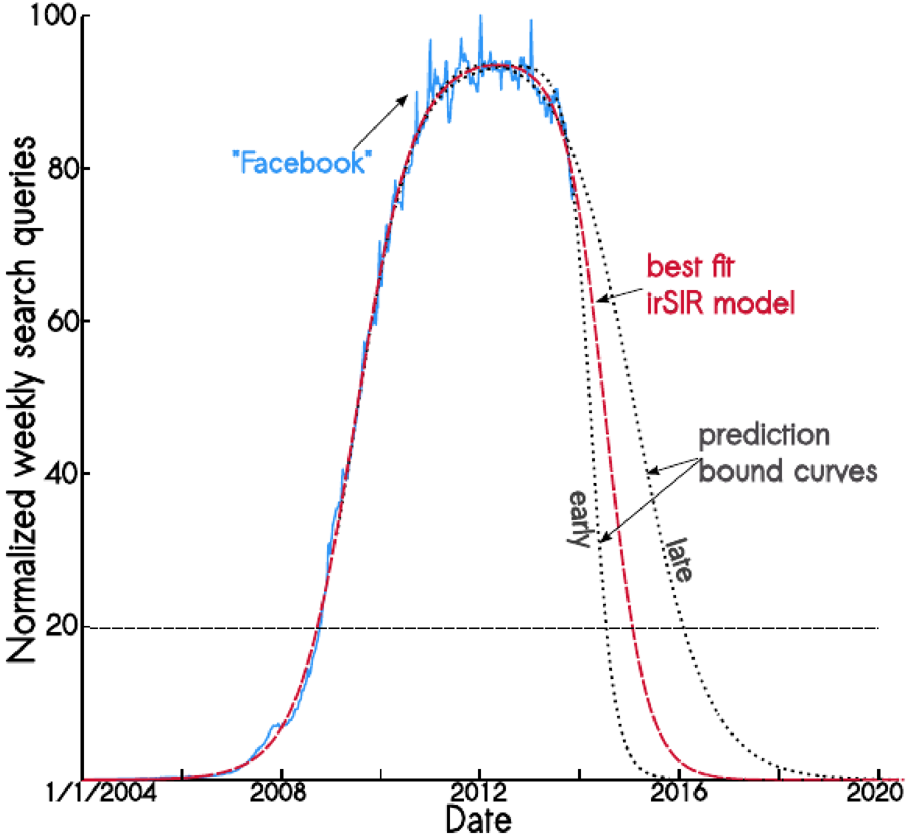


Backtesting with Friendster's collapse



Constantly decreasing k_s in simple k-core model. Inset: fraction of nodes with coreness below the median over the lifetime of Friendster.

Predicting Facebook's collapse

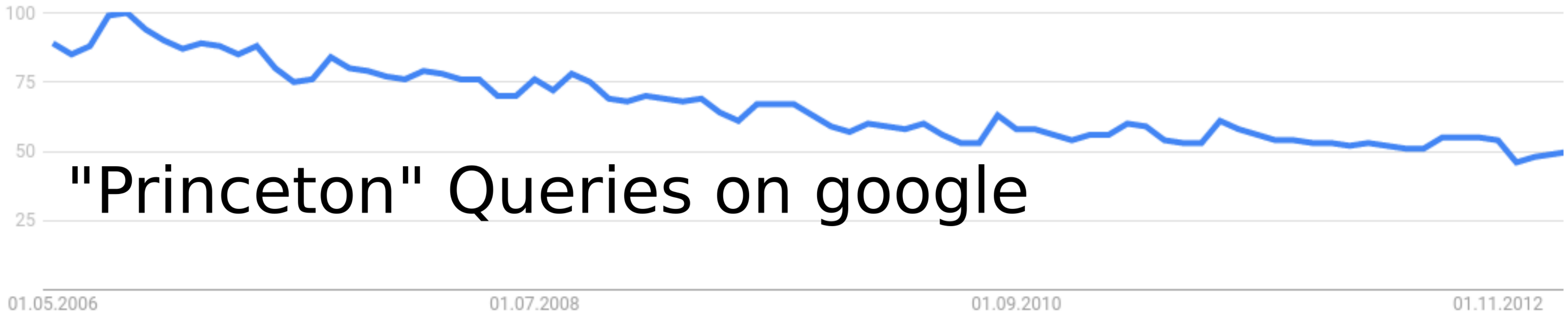


- Cannarella and Spechler, 2014
- Epidemics model applied to the decay of online social networks
- Google trends to measure the number of active Facebook users
- Extrapolation predicted that Facebook would lose 80% of its users by 2017

Epidemiological modeling of online social network dynamics. John Cannarella, Joshua A. Spechler. Arxiv preprint (2014)

Stretching the Google trends method

Data scientists at Facebook replied to the Arxiv paper showing the problem with measuring social network use levels using Google Trends data. Applying the same methodology, Facebook researchers reached the conclusion that Princeton would lose 80% of its students by 2021:

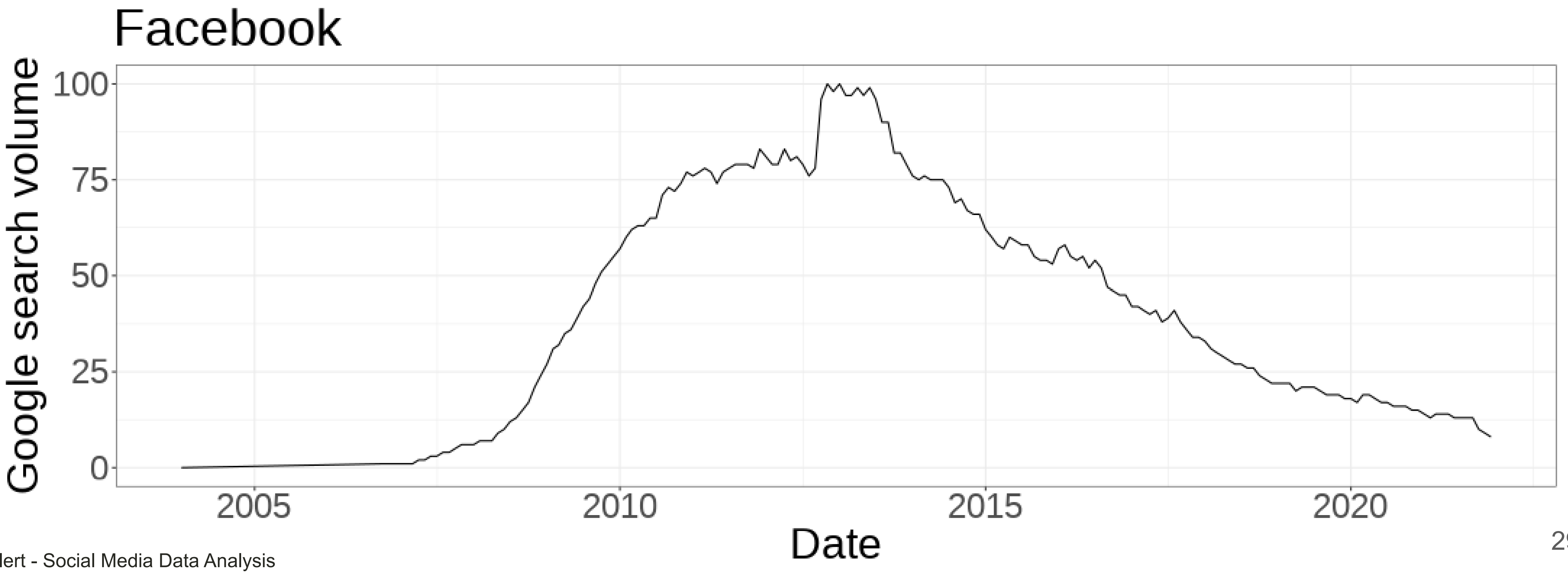


Stretching the Google trends method

You can even apply it to air and come to the conclusion that the atmosphere will run out of air by 2060:

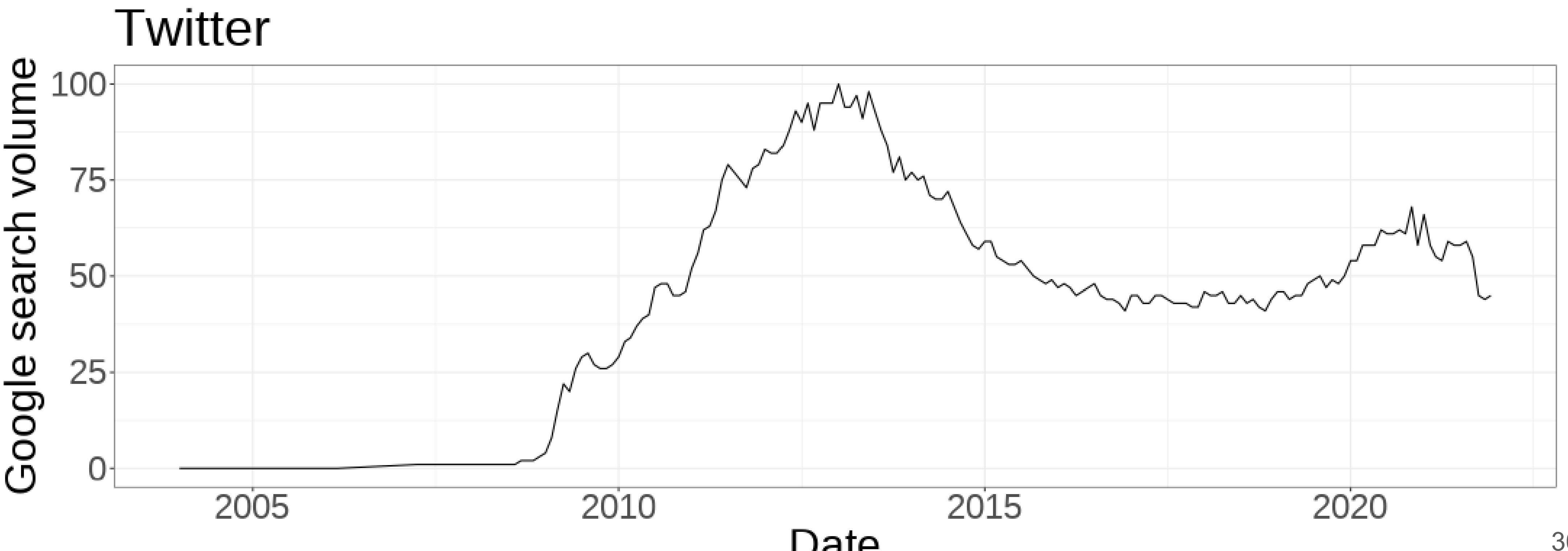
Google trends for Facebook

The Facebook examples show that decrease in search volume is a decrease in information searching about the social network, not a decrease in access and use.



Google trends for Twitter

Twitter is not living a second growth, in fact it's growth has been rather slow to stagnant for a few years, even though it appears it is gaining users after Trump's permanent suspension.



Temporal validity issues

A more accurate way to measure activity in a social network is Bruno Ribeiro's approach using Alexa data, but Alexa focuses on website visits without considering access through mobile apps.

Measurement is always an important issue in Computational Social Systems, and just because a paper used a measurement method few years ago, it does not mean it is valid today.

> **Take home message:** Your measures based on today's digital traces might not work on tomorrow's

Summary

- Measuring importance as centrality in social networks
  - Degree: when having lots of links makes you important
  - Betweenness: when being in shorter paths makes you important
  - Closeness: When having short distances to the rest makes you important

- Social resilience as networked decisions
  - Modeling human decisions rather than disease spreading
  - A decision of a user leaving affects the situation of other users
  - This can be mapped to network metrics like coreness centrality

- Empirical analysis of online social resilience
  - Using trends as activity and sometimes social network data
  - Historical analysis of the Friendster collapse
  - Limits of Google trends: the Facebook case