Social Networks: Information Cascades and Social Influence Maximization

Advanced Computer Networks

Summer Semester 2013

Recap: How Networks Form?

- Random Network
- Small-world Network
- Scale-free Network

Recap: The Random Network

- Erdös-Renyi Random Graph Model
  - Fix a set of n nodes, N={1,2,…,n}.
  - Each link is formed with a given probability p (0<p<1), and the formation of links is independent.
- Degree distribution of random graph: approximated by a Poisson distribution

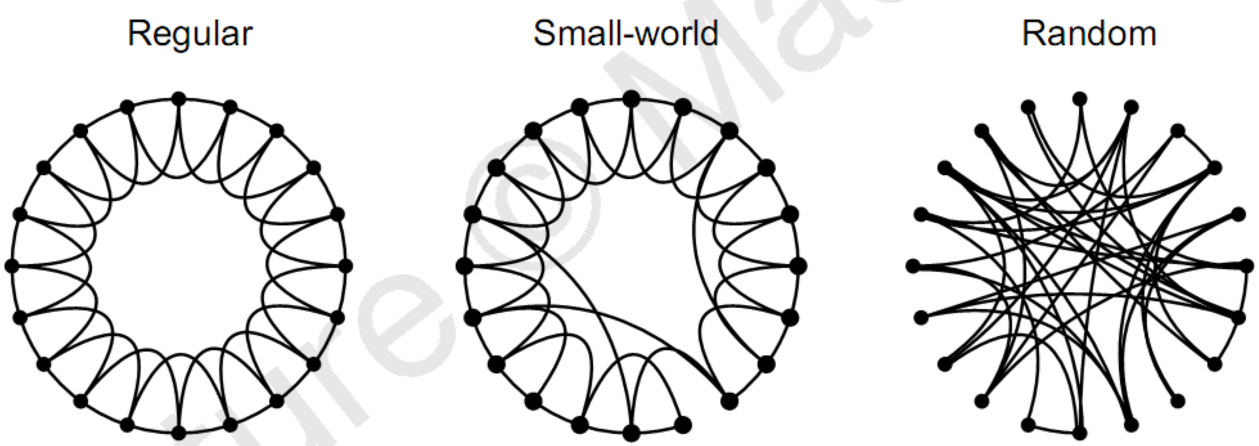
$$\Pr(d) \approx \frac{e^{-(n-1)p}((n-1)p)^d}{d!}$$

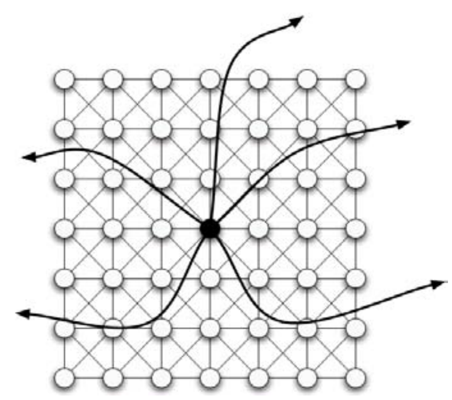
- Threshold for phase transition:

  p>log(n)/(n-1)

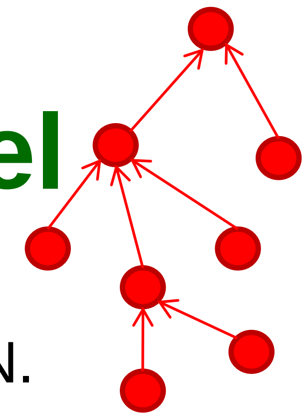
Recap: The Small-World Model

- Can we make up a simple model that exhibits both of the features: many closed triads, but also very short paths?
- **One-deminsional Model (Watts-Strogatz)**
- Starting from a ring lattice with n vertices and k edges per vertex.
  - Regular network with high clustering coefficient
- We rewire each edge at random with probability p (0≤p≤1).
  - p=0: regular network
  - p=1: random network
  - Randomizing the network, lowering average path length





Recap: Rich Get Richer Model



- Creation of links among Web pages
  - Pages are created in order, and named 1; 2; 3; …;N.
  - When page j is created, it produces a link to an earlier Web page i according to:
  - 1) With prob. p (0<p<1), j links to i chosen uniformly at random (from among all earlier nodes)
  - 2) With prob. 1-p, node j links to node u with prob. proportional to the degree of u
- Major results: let q=1-p, for degree k, by estimation

$$Pr\{x \geq k\} = [\frac{q}{p} \cdot k + 1]^{-1/q}$$

$$F\{x\} = Pr\{x < k\} = 1 - Pr\{x \geq k\}$$

$$Pr\{x = k\} = F'(x) = \frac{1}{p}[\frac{q}{p} \cdot k + 1]^{-(1+1/q)}$$

Power-Law!

Information Cascades

- Examples
  - Choosing the side in a war
    - Side A 70% chance to win
    - Side B 30% chance to win
  - Choosing a restaurant in an unfamiliar town
    - Restaurant A 0 guests
    - Restaurant B 100 guest
    - Your private information: A received good comments in Internet
  - Looking into the sky
  - New products, ideas, …
- Influence of human behaviors and decisions
  - Following the crowd

Information Cascade

- An information cascade may occur when people make decisions sequentially
  - Later people watching the actions of earlier people, and from these actions inferring something about what the earlier people know.
  - In the restaurant example, when the first diners chose restaurant B, they conveyed information to later diners.
  - A cascade then develops when people abandon their own information in favor of inferences based on earlier people's actions.
- Individuals in a cascade are imitating the behavior of others, but it is from rational inferences of limited information

Milgram et.al. [1969]

- Groups of people (from 1-15 people) stand on a street corner and stare up into the sky
- What happen to the passersby?
  - If one person looking up, very few passersby stopped.
  - If 5 person looking up, more passersby stopped, but most still ignored them
  - If 15 people looking up, 45% of passersby stopped and also stared up into the sky
- Reference: Stanley Milgram, Leonard Bickman, and Lawrence Berkowitz. Note on the drawing power of crowds of dierent size. Journal of Personality and Social Psychology, 13(2):79{82, October 1969.

Basic Ingredients

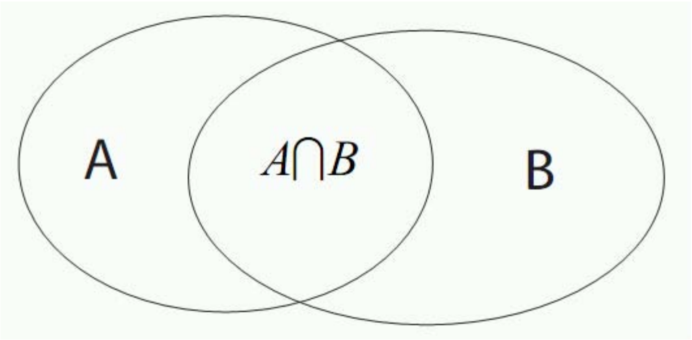
- Herding (Information Cascade)
  - There is a decision to be made
  - People make the decision sequentially
  - Each person has some private information that helps guide the decision
  - You can't directly observe private information of the others but can see what they do

A Simple Example

- Consider an urn with 3 marbles. It can be either:
  - Majority-blue: 2 blue, 1 red
  - Majority-red: 1 blue, 2 red
- Each person wants to best guess whether the urn is majority-blue or majority-red
- Experiment: One by one each person:
  - Draws a marble
  - Privately looks are the color and puts the marble back
  - Publicly guesses whether the urn is majority-red or majority-blue
- You see all the guesses beforehand.
- How should you make your guess?

- What happens?
  - #1 person: Guess the color you draw from the urn.
  - #2 person: Guess the color you draw from the urn.
    - If same color as 1st, then go with it
    - If different, break the tie by doing with your own color
  - #3 person:
    - If the two before made different guesses, go with your color
    - Else, go with their guess (regardless your color) – cascade starts!
  - #4 person:
    - Suppose the first two guesses BLUE, you go with BLUE (Since 3rd person always guesses BLUE)
  - Everyone else guesses BLUE (regardless of their draw)

Revisit Probabilistic Theory



- Conditional probability of A given B

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

- Similarly

$$\Pr[B \mid A] = \frac{\Pr[B \cap A]}{\Pr[A]}$$

- Rewriting as: $\Pr[A \mid B] \cdot \Pr[B] = \Pr[A \cap B] = \Pr[B \mid A] \cdot \Pr[A]$
- Bayes' Rule (posterior probability):

$$\Pr[A \mid B] = \frac{\Pr[A] \cdot \Pr[B \mid A]}{\Pr[B]}.$$

Analysis

Bayes' Rule (posterior probability):

$$\Pr[A \mid B] = \frac{\Pr[A] \cdot \Pr[B \mid A]}{\Pr[B]}$$

- #1 follows her own color (private signal)
  - Prior probabilities $\Pr[\mathit{majority\text{-}blue}] = \Pr[\mathit{majority\text{-}red}] = \frac{1}{2}.$
  - For the two kinds of urns

$$\Pr[\mathit{blue} \mid \mathit{majority\text{-}blue}] = \Pr[\mathit{red} \mid \mathit{majority\text{-}red}] = \frac{2}{3}.$$

  - If a blue marble is drawn

$$\Pr[\mathit{majority\text{-}blue} \mid \mathit{blue}] = \frac{\Pr[\mathit{majority\text{-}blue}] \cdot \Pr[\mathit{blue} \mid \mathit{majority\text{-}blue}]}{\Pr[\mathit{blue}]}.$$

$$\begin{aligned}\Pr[\mathit{blue}] &= \Pr[\mathit{majority\text{-}blue}] \cdot \Pr[\mathit{blue} \mid \mathit{majority\text{-}blue}] + \\ &\quad \Pr[\mathit{majority\text{-}red}] \cdot \Pr[\mathit{blue} \mid \mathit{majority\text{-}red}] \\ &= \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}.\end{aligned}$$

  - Thus

$$Pr[majority - blue|blue] = \frac{\frac{1}{2} \cdot \frac{2}{3}}{\frac{1}{2}} = \frac{2}{3}$$

- #2 student
  - #2 knows #1's color. So, #2 gets 2 colors.
    - If they are the same, decision is easy.
    - If not, break the tie in favor of her own color
- #3 follows majority signal
  - Knows #1, #2 acted on their colors. So, #3 gets 3 signals.
    - If #1 and #2 made opposite decisions, #3 goes with her own color.
    - If #1 and #2 made same choice, #3 follows then (Her decision conveyed no info. Cascade has started!)

$$Pr[majority - blue|blue, blue, red] = ?$$

$$Pr[majority - blue|blue, blue, red] =?$$

- According to Bayes' Rule

$$\Pr\left[majority\text{-}blue \mid blue,\ blue,\ red\right] = \frac{\Pr\left[majority\text{-}blue\right] \cdot \Pr\left[blue,\ blue,\ red \mid majority\text{-}blue\right]}{\Pr\left[blue,\ blue,\ red\right]}.$$

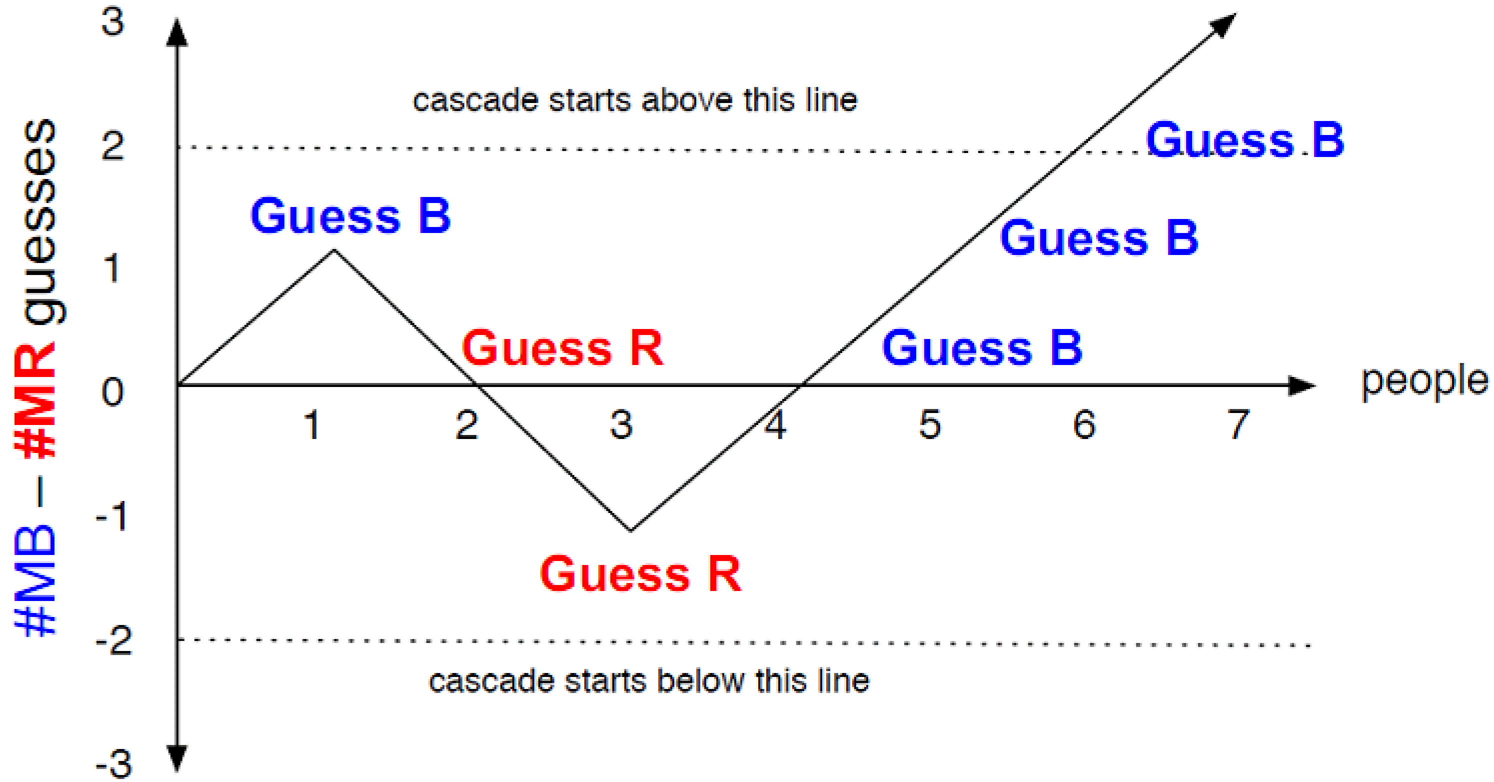
- Since

$$\Pr\left[blue,\ blue,\ red \mid majority\text{-}blue\right] = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{4}{27}.$$

$$\begin{aligned}\Pr\left[blue,\ blue,\ red\right] &= \Pr\left[majority\text{-}blue\right] \cdot \Pr\left[blue,\ blue,\ red \mid majority\text{-}blue\right] + \\ &\quad \Pr\left[majority\text{-}red\right] \cdot \Pr\left[blue,\ blue,\ red \mid majority\text{-}red\right] \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{6}{54} = \frac{1}{9}.\end{aligned}$$

- So

$$\Pr\left[majority\text{-}blue \mid blue,\ blue,\ red\right] = \frac{\frac{4}{27} \cdot \frac{1}{2}}{\frac{1}{9}} = \frac{2}{3}.$$



#MB – #MR guesses
3
2
1
0
-1
-2
-3
cascade starts above this line
cascade starts below this line
Guess B
Guess R
Guess R
Guess B
Guess B
Guess B
people
1
2
3
4
5
6
7

Lessons from Cascades

- Cascades can be wrong
  - A cascade of acceptances will start when the first two people happen to get high signals, even though it is the wrong choice for the population
- Cascades can be based on very little information.
  - Since people ignore their private information once a cascade starts, only the pre-cascade information influences the behavior of the population.
- Cascades can be fragile
  - Suppose the first two guess blue
  - Student #1 to #100 guess blue
  - If student #99 and #100 draw red and show then in public
  - Student #101 now has 4 pieces of information, and she guesses based on her own color
  - Cascade is broken!

computer NETWORKS

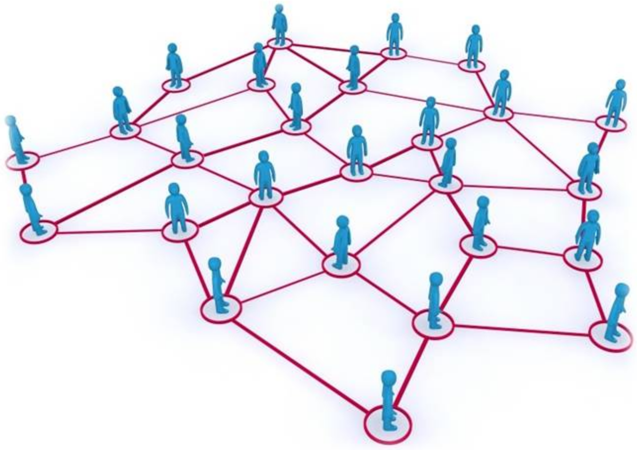
Summary

- Information cascades and rich get richer may explain many social behaviors
  - Rumors
  - New technology
  - Fashions
  - Keeping your money or not in a stock market
  - Voting for popular candidates
  - Best selling books, music
  - Riots, protests, strikes

Social Influence Maximization

Social Networks

- Nodes: people (*actors*)
- Edges: social interactions, friendship, etc. (*ties*)
- Not to be confused with Online Social Network sites, which are merely representations

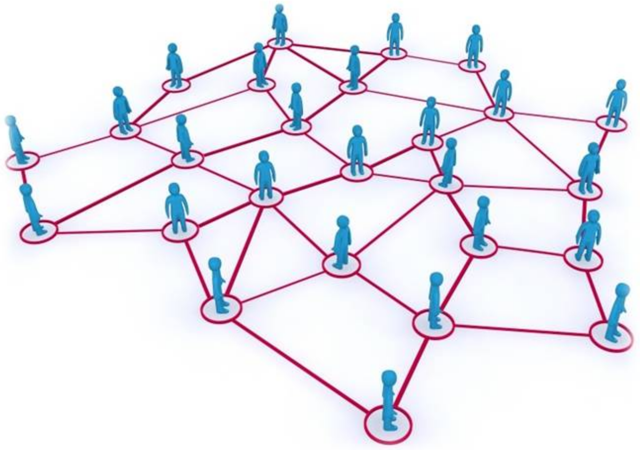


Social Influence

- People's behaviour is affected by other people's behaviour
- Information: „*Many of my friends are using Linux, so there must be some advantage in that.*“
- Externalities: „*Many of my friends are using XMPP/Jabber for instant messaging, so it is of much more use to me than, say, Skype.*“
- Trends: „*Prestigious people around me have started to wear red shoes, so it must be trendy.*“

Who is the most influential?

- Natural question in society
- Interesting for understanding processes like the spread of behaviour, rumours, collective dynamics, etc.
- Useful for viral marketing: targeted advertising



Modelling with Probabilities

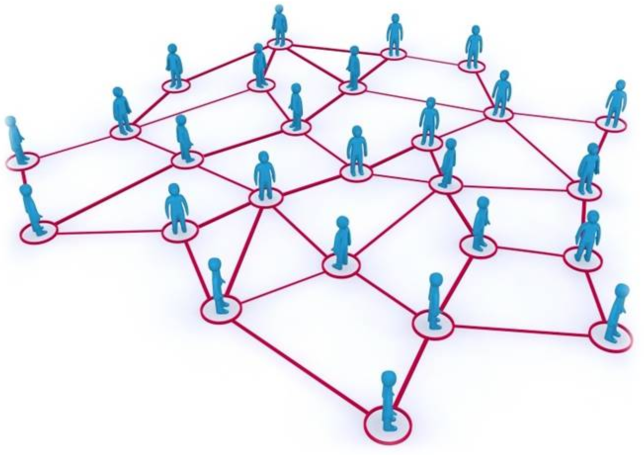
- Random variables:
  X_v = (Person v adopts ? 1 : 0)
- Assumed to build a Markov field:
  $P(X_v \mid X_w$, w *anybody else*)
  = $P(X_v \mid X_w$, w *connected to v*)

Difficulties

- Very general model
- Extremely many parameters:
  exponentially many conditional probabilities
- We need some simplifying assumptions to get a simpler model.

Deterministic threshold models

- Every node (person) has some threshold between 0 and 1
- A node adopts the new behaviour once the fraction of their neighbours who have already adopted it exceeds the given threshold
- Background motivation: simple game-theoretic approach

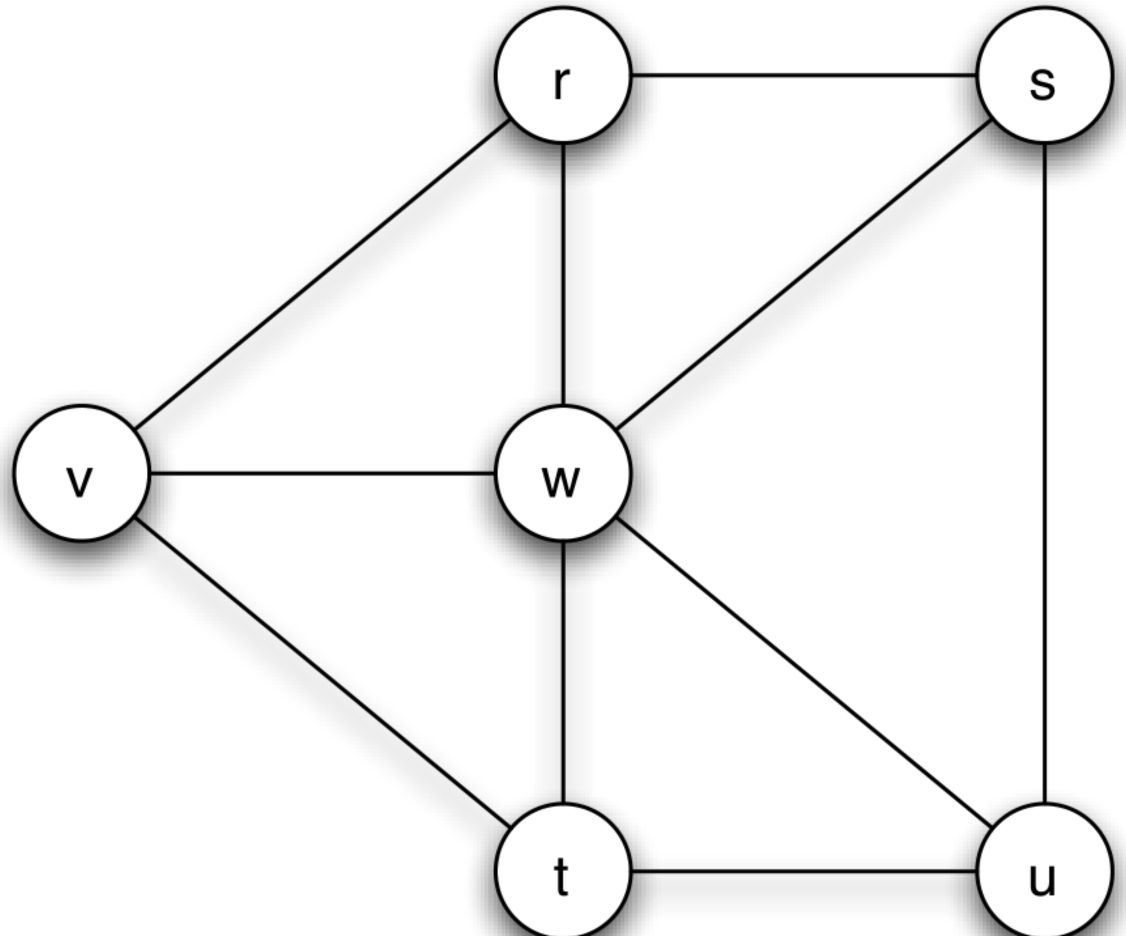


Emergence of thresholds

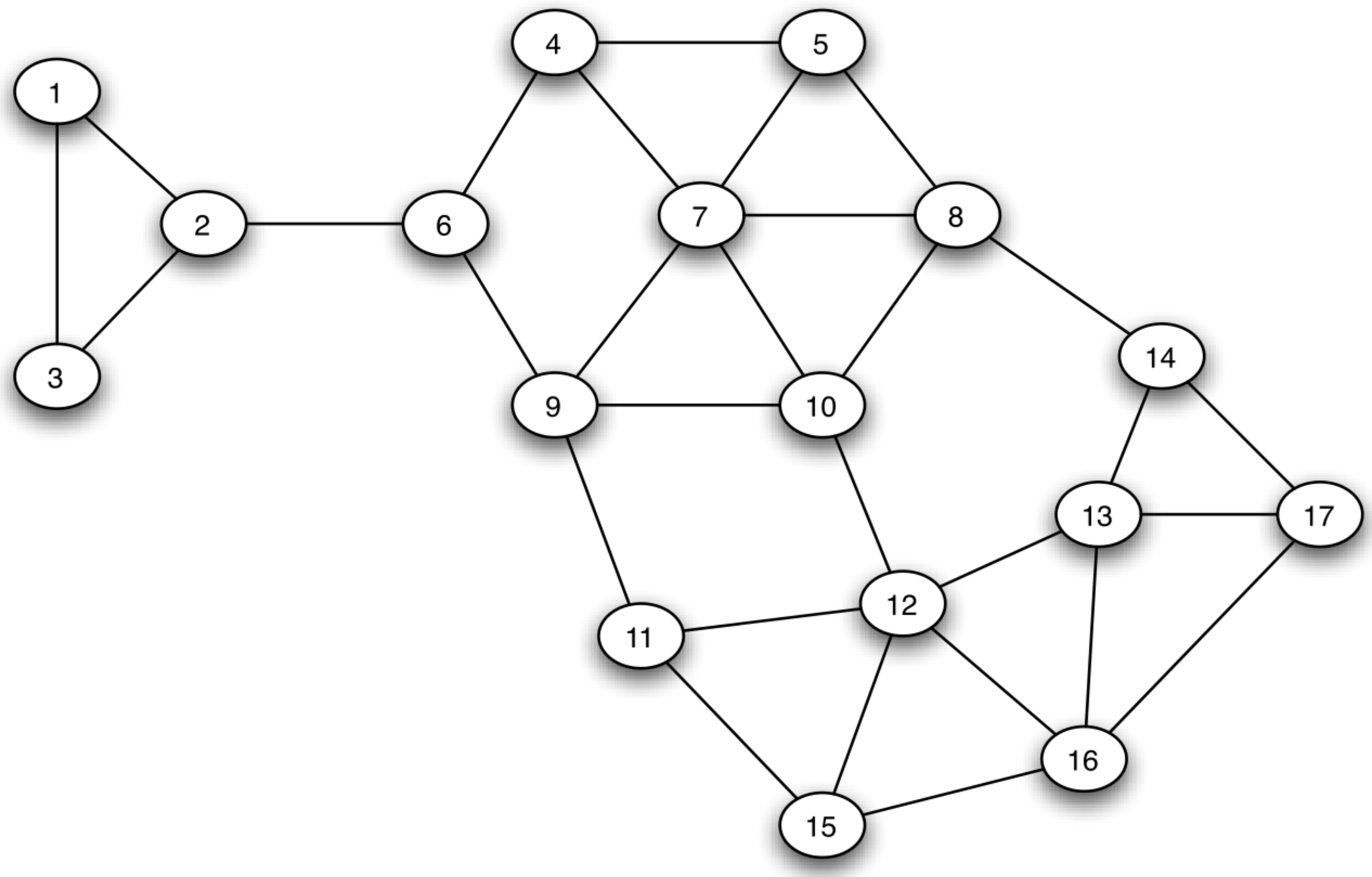
- Assume there is an old behaviour A and a new behaviour B where everybody can choose one
- Every person gets a certain payoff from each relationship in which their own chosen behaviour matches the respective neighbour's
- If both choose A, each of them gets payoff a;
  if both choose B, each of them gets payoff b
- An individual's payoffs accumulate over all of their neighbours

Calculation

Example 1

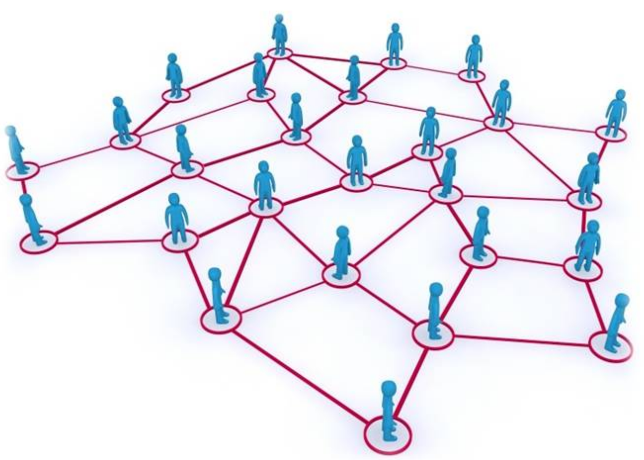


Example 2



Maximization problem

- Given a network (nodes and edges), node thresholds and weights on edges indicating the strength of the relationship
- Assume you can select k nodes in the network that will initially adopt the new behaviour
- Which k nodes should you choose in order to maximize the number of nodes eventually adopting the new behaviour?

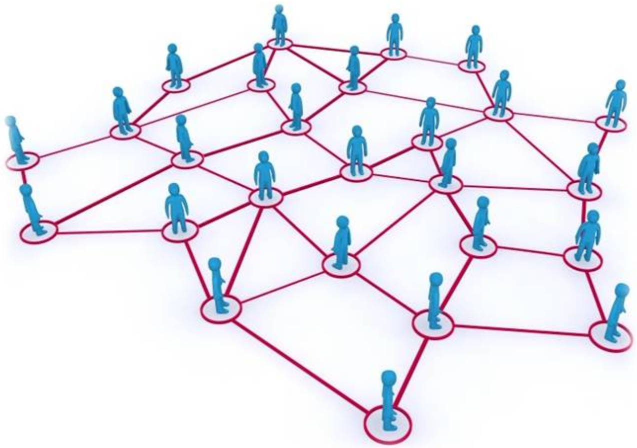


Maximization is NP-hard

- Even in the case of unweighted edges and a global threshold value, this problem includes the NP-complete VERTEX-COVER problem as special case.

Probabilistic influence models

- Whenever a node adopts the new behaviour, they have the one-time chance to *infect* each of their neighbours who have not yet adopted it
- The probability that they succeed is given by the weight of the corresponding edge
- Now the expected number of nodes eventually adopting the new behaviour is to be maximized

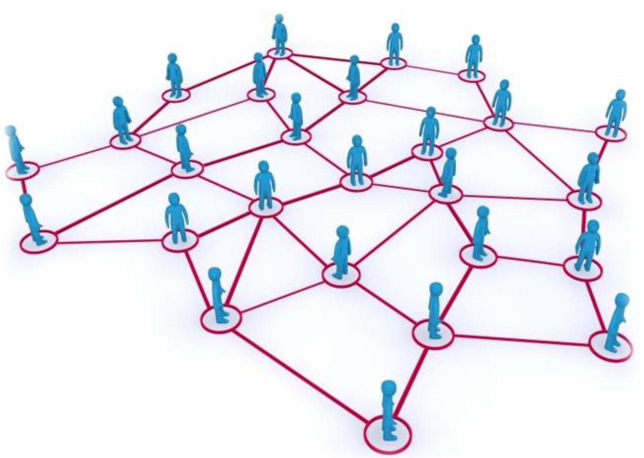


Maximization is NP-hard

- This problem includes the NP-complete SET-COVER problem as special case.

Heuristics for Approximation

- Node degree: select the nodes with the highest numbers of neighbours (friends)
- Distance centrality: select the nodes with the lowest average distance to all others
- Greedy hill-climbing algorithm: in each step, select the node for which the objective grows the most



Comparison of Heuristics

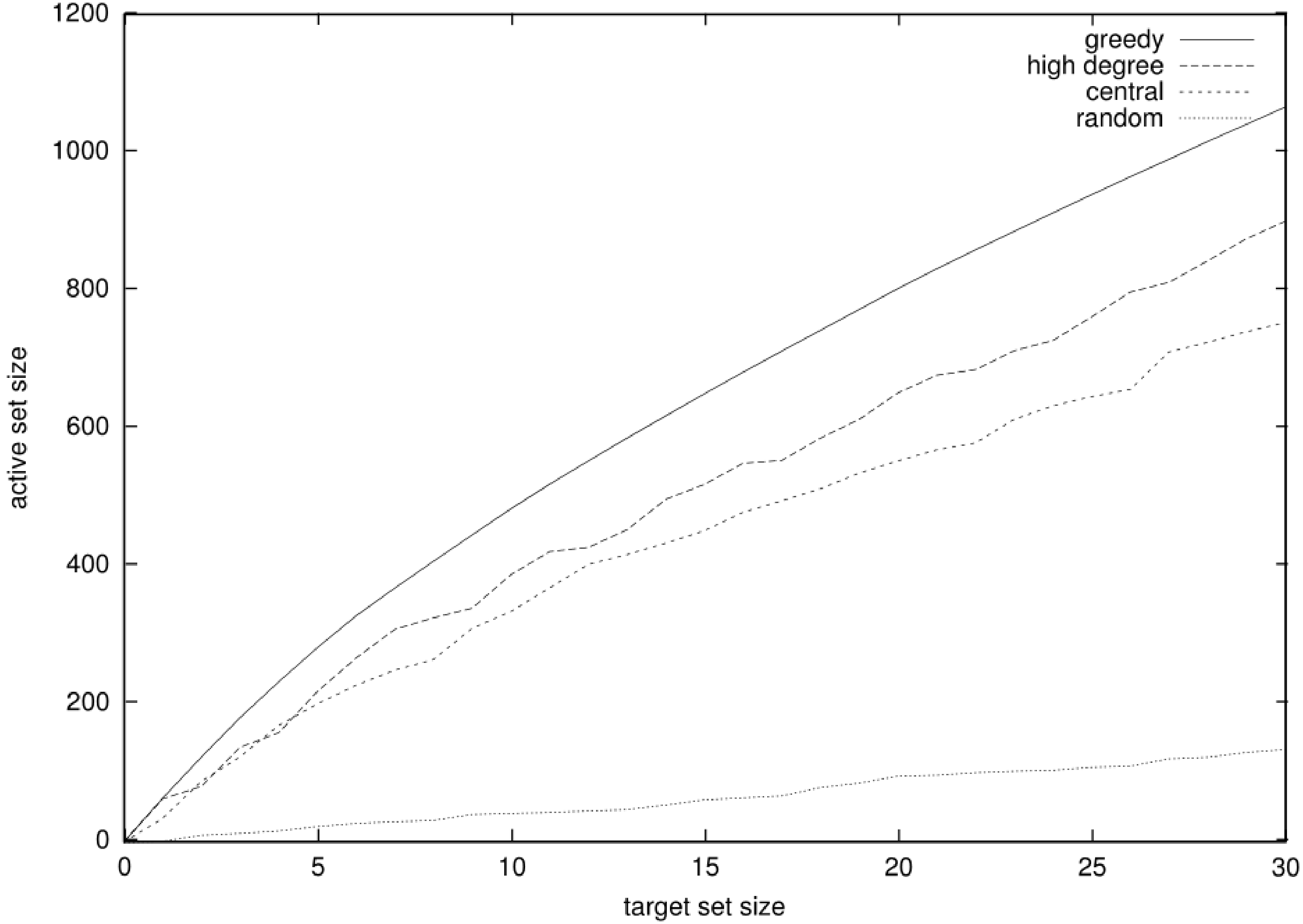


Figure 1: Results for the linear threshold model

Comparison of Heuristics

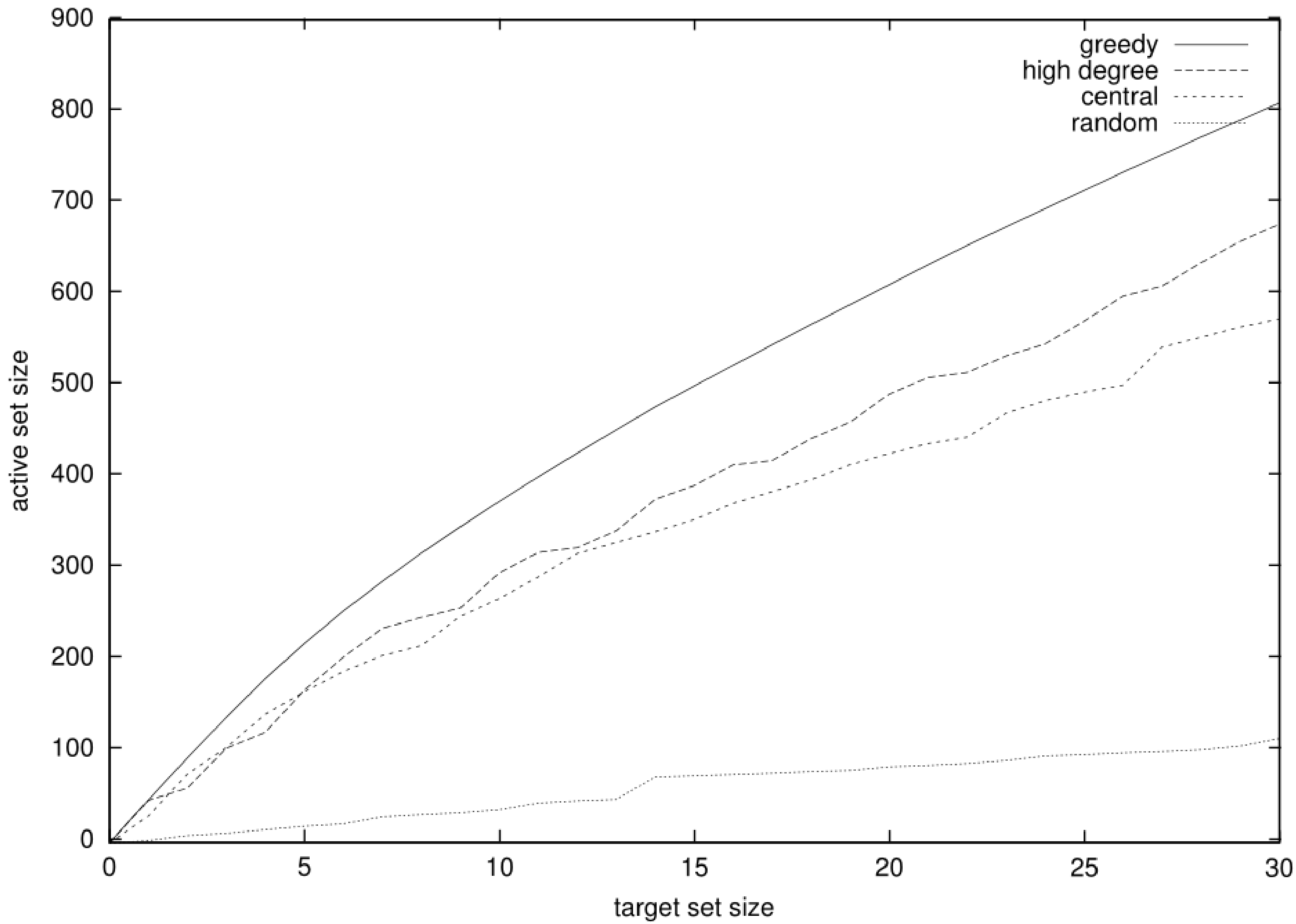


Figure 2: Results for the weighted cascade model

Comparison of Heuristics

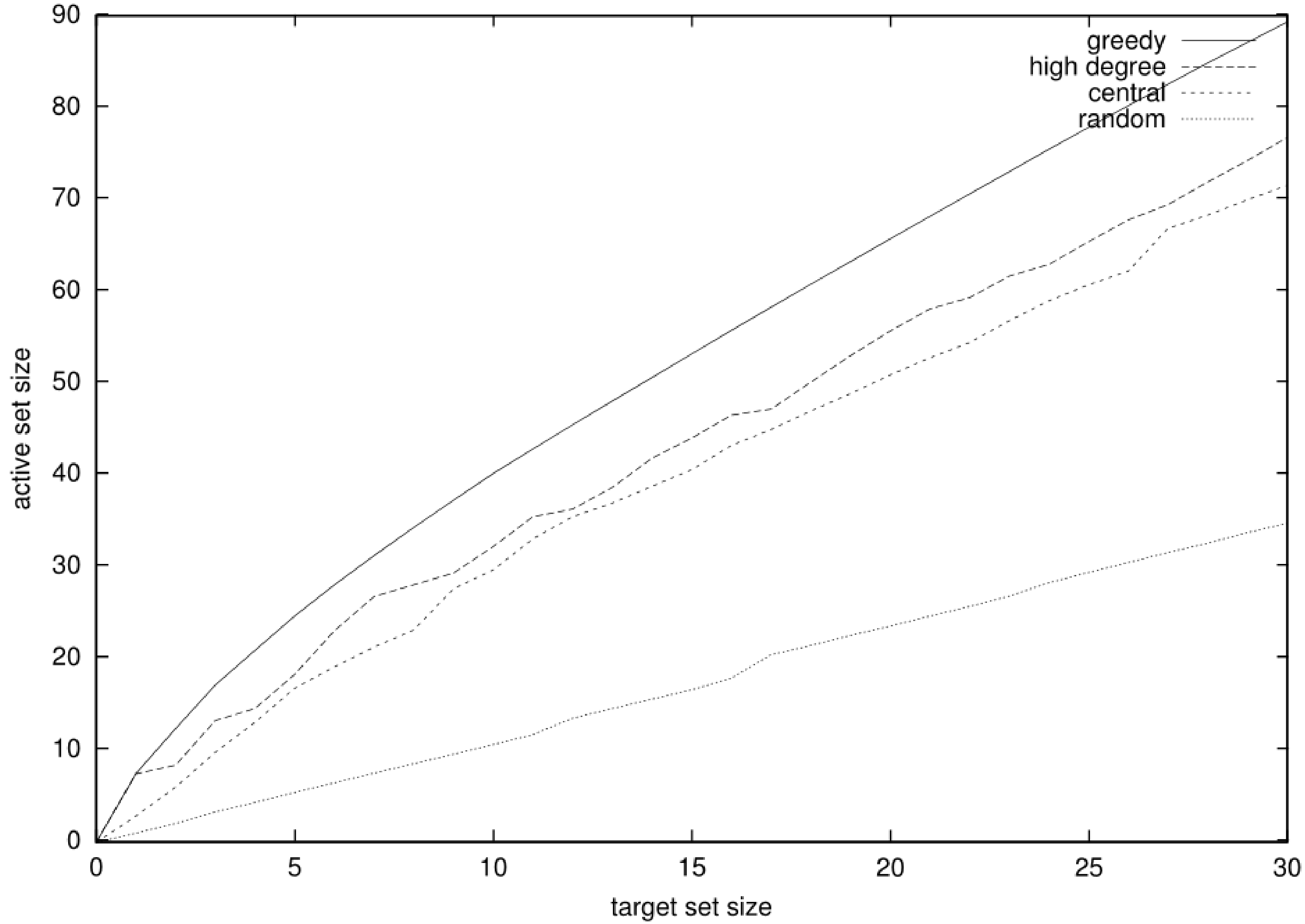


Figure 3: Independent cascade model with probability 1%

Comparison of Heuristics

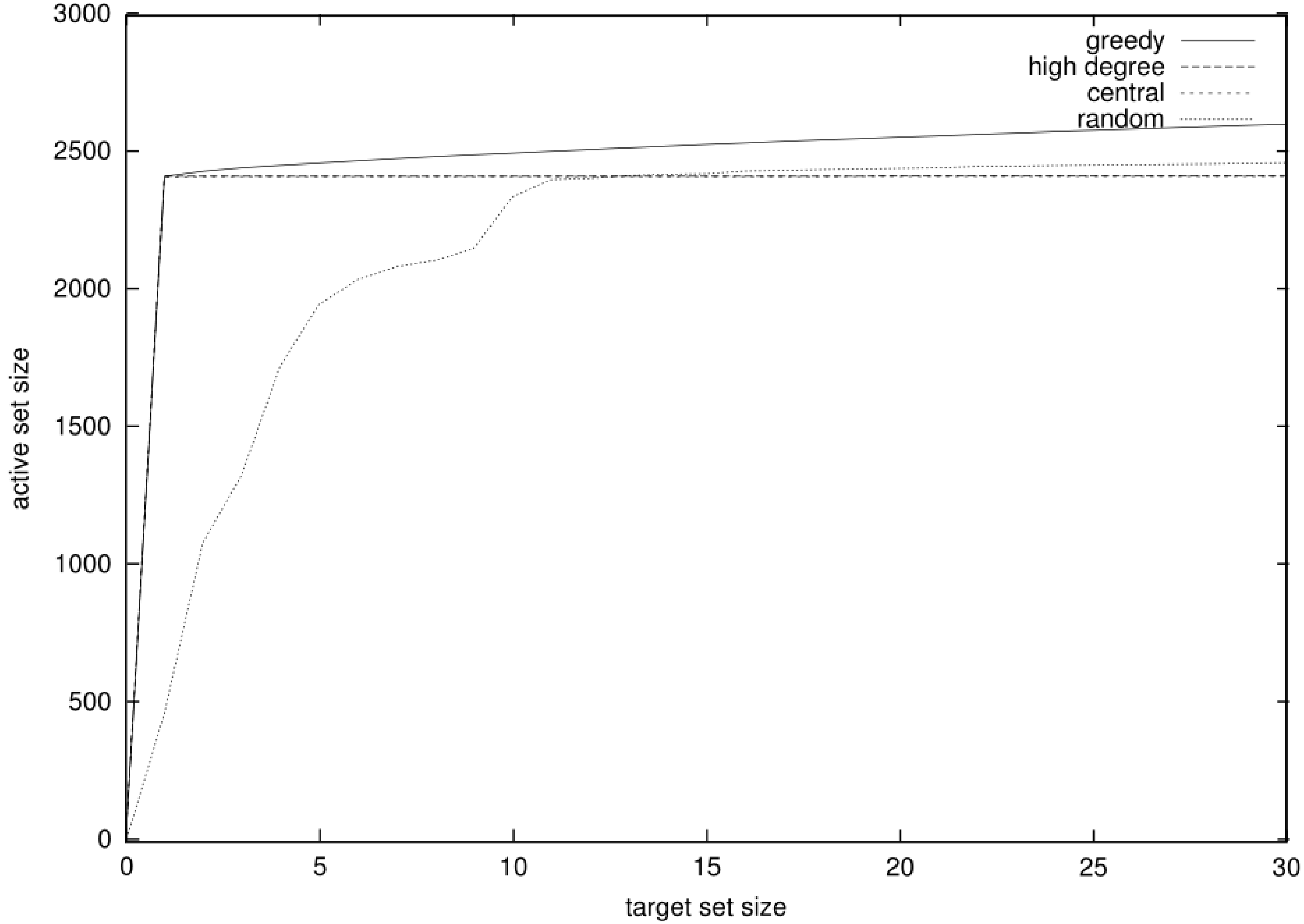


Figure 4: Independent cascade model with probability 10%