Online Social Networks and Media

Opinion formation on social networks

Diffusion of items

- So far we have assumed that what is being diffused in the network is some discrete item:
  - E.g., a virus, a product, a video, an image, a link etc.
- For each network user a binary decision is being made about the item being diffused
  - Being infected by the virus, adopt the product, watch the video, save the image, retweet the link, etc.
  - This decision may happen with some probability, but the probability is over the discrete values {0,1} and the decisions usually do not change

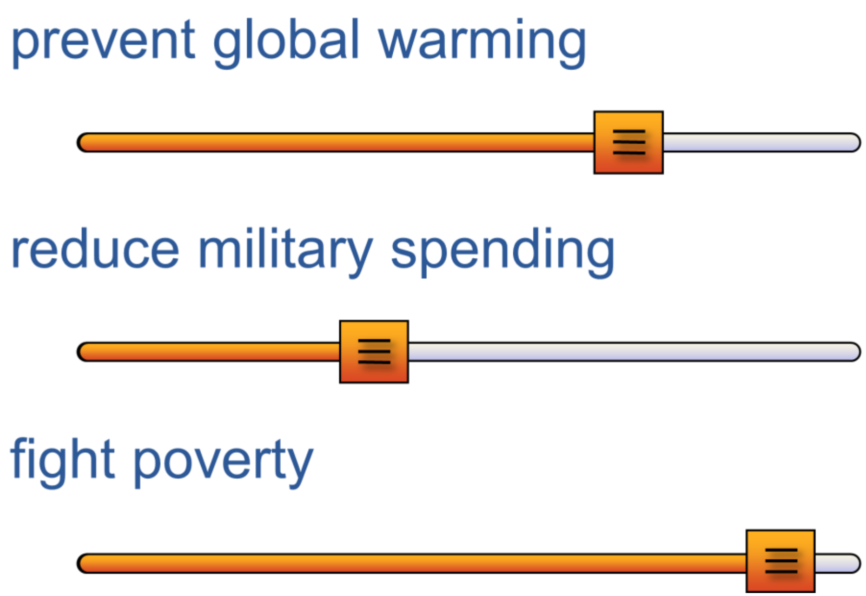
Diffusion of opinions

- The network can also diffuse opinions.
  - What people believe about an issue, a person, an item, is shaped by their social network
- People hold opinions that may change over time due to social influence
- Opinions may assume a continuous range of values, from completely negative to completely positive.
  - Opinion diffusion is different from item diffusion
  - It is often referred to as opinion formation.
- The most similar case we have seen is the game-theoretic cascades

What is an opinion?

- An opinion is a real value
  - In our models a value in the interval [0,1]

(0: negative, 1: positive)



How are opinions formed?

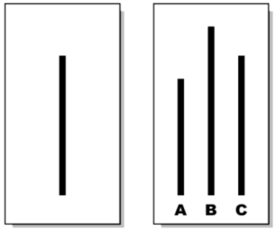
- Opinions change over time

How are opinions formed?

- And they are influenced by our social network

Social Influence

- There are two main types of social influence:
  - Normative Influence: Users influenced by opinion of neighbors due to social norms, conformity, group acceptance, avoiding ridicule, etc
  - Informational Influence: Users lacking necessary information, or not trusting their information, use opinion of neighbors to form their opinions
- Asch's conformity experiment [55]:



Modeling opinion formation

- There is a lot of work from different perspectives:
  - Psychologists/Sociologists: field experiments and decades of observations
  - Statistical Physicists: model humans as particles and predict their behavior
  - Mathematicians/Economists: Use game theory to model human behavior
  - Computer Scientists: build algorithms on top of the models
- Questions asked:
  - How do societies reach consensus?
    - Not always the case, but necessary for many issues in order for society to function
  - When do we get polarization or opinion clusters?
    - More realistic in the real world where consensus tends to be local

De Groot opinion formation model

- Every user i has an opinion $z_i \in [0,1]$
- The opinion of each user in the network is iteratively updated, each time taking the average of the opinions of its neighbors and herself

$$z_i^t = \frac{w_{ii} z_i^{t-1} + \sum_{j \in N(i)} w_{ij} z_j^{t-1}}{w_{ii} + \sum_{j \in N(i)} w_{ij}}$$

  - where $N(i)$ is the set of neighbors of user i.
- This iterative process converges to a consensus

What about personal biases?

- People tend to cling on to their personal opinions

Another opinion formation model (Friedkin and Johnsen)

- Every user i has an intrinsic opinion $s_i \in [0,1]$ and an expressed opinion $z_i \in [0,1]$
- The public opinion z_i of each user in the network is iteratively updated, each time taking the average of the expressed opinions of its neighbors and the intrinsic opinion of herself

$$z_i^t = \frac{w_{ii} s_i + \sum_{j \in N(i)} w_{ij} z_j^{t-1}}{w_{ij} + \sum_{j \in N(i)} w_{ij}}$$

Opinion formation as a game

- Assume that network users are rational (selfish) agents
- Each user has a personal cost for expressing an opinion

$$c(z_i) = w_{ii}(z_i - s_i)^2 + \sum_{j \in N(i)} w_{ij}(z_i - z_j)^2$$

Inconsistency cost: The cost for deviating from one's intrinsic opinion

Conflict cost: The cost for disagreeing with the opinions in one's social network

- Each user is selfishly trying to minimize her personal cost.

D. Bindel, J. Kleinberg, S. Oren. *How Bad is Forming Your Own Opinion?* Proc. 52nd IEEE Symposium on Foundations of Computer Science, 2011.

Opinion formation as a game

- The opinion z_i that minimizes the personal cost of user i

$$z_i = \frac{w_{ij} s_i + \sum_{j \in N(i)} w_{ij} z_j}{w_{ij} + \sum_{j \in N(i)} w_{ij}}$$

- In linear algebra terms (assume 0/1 weights):

$$(L + I)\boldsymbol{z} = \boldsymbol{s} \;\Rightarrow \boldsymbol{z} = (L + I)^{-1}\boldsymbol{s}$$

where L is the Laplacian of the graph.

Reminder: The Laplacian is the negated adjacency matrix with the degree on the diagonal

Understanding opinion formation

- To better study the opinion formation process we will show a connection between opinion formation and absorbing random walks.

Random Walks on Graphs

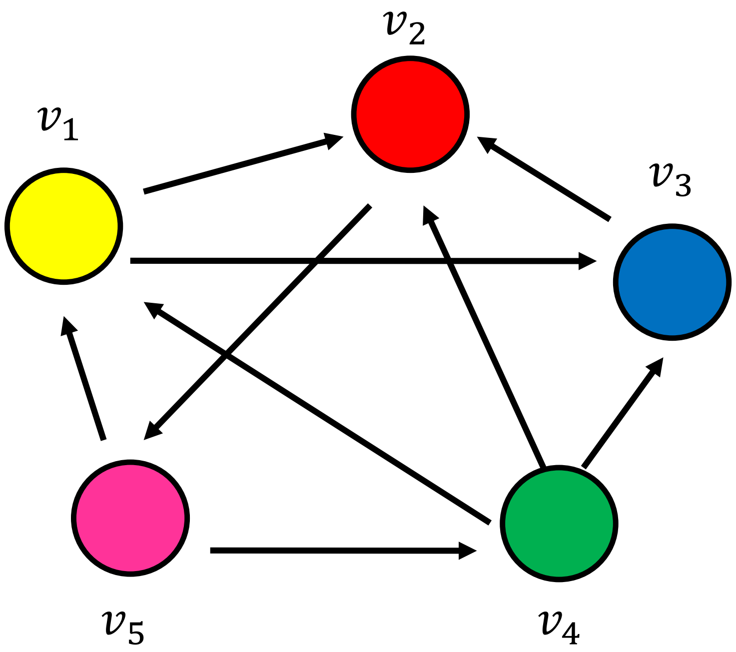
- A random walk is a stochastic process performed on a graph

- Random walk:
  - Start from a node chosen uniformly at random with probability $\frac{1}{n}$.
  - Pick one of the outgoing edges uniformly at random
  - Move to the destination of the edge
  - Repeat.

- Made very popular with Google's PageRank algorithm.

The Transition Probability matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$



$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

$P[i,j] = 1/d_{out}(i)$: Probability of transitioning from node i to node j.

Random walk

- Question: what is the probability p_i^t of being at node i after t steps?

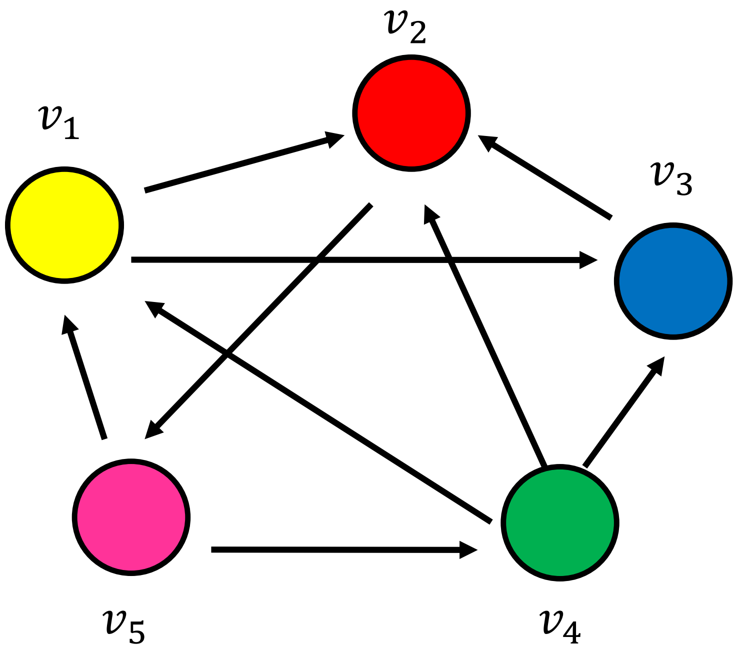
$$p_1^0 = \frac{1}{5} \qquad p_1^t = \frac{1}{3} p_4^{t-1} + \frac{1}{2} p_5^{t-1}$$

$$p_2^0 = \frac{1}{5} \qquad p_2^t = \frac{1}{2} p_1^{t-1} + p_3^{t-1} + \frac{1}{3} p_4^{t-1}$$

$$p_3^0 = \frac{1}{5} \qquad p_3^t = \frac{1}{2} p_1^{t-1} + \frac{1}{3} p_4^{t-1}$$

$$p_4^0 = \frac{1}{5} \qquad p_4^t = \frac{1}{2} p_5^{t-1}$$

$$p_5^0 = \frac{1}{5} \qquad p_5^t = p_2^{t-1}$$



$$p^t = p^{t-1} P$$

Node Probability vector

- The vector $p^t = (p_1^t, p_2^t, \dots, p_n^t)$ that stores the probability of being at node v_i at step t
  - p_i^0 = the probability of starting from state i (usually) set to uniform
- We can compute the vector p^t at step t using a vector-matrix multiplication

$$p^t = p^{t-1} P = p^0 P^t$$

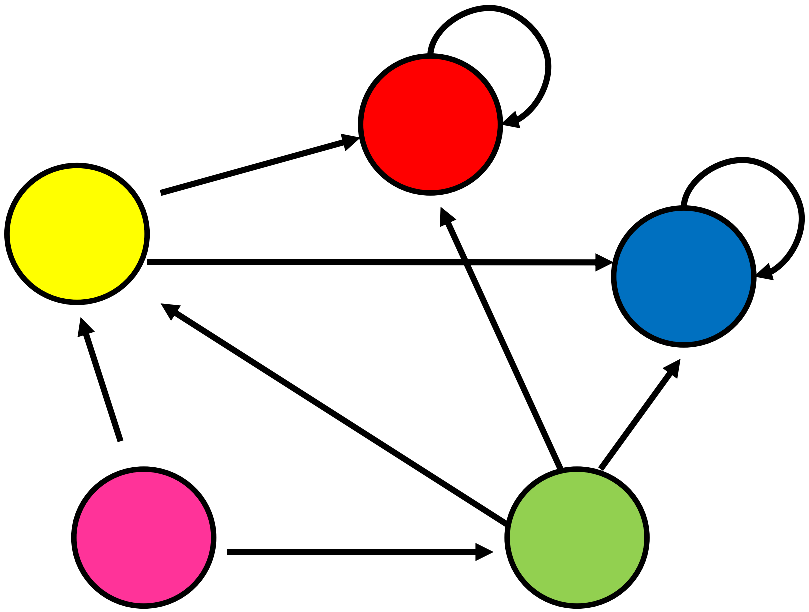
- After many steps $p^t \to \pi$ the probability converges to the stationary distribution π

Stationary distribution

- The stationary distribution of a random walk with transition matrix P, is a probability distribution π, such that $\pi = \pi P$
- The stationary distribution is independent of the initial vector if the graph is strongly connected, and not bipartite.
- All the rows of the matrix P^{∞} are equal to the stationary distribution π
- The stationary distribution is an eigenvector of matrix P
  - the principal left eigenvector of P – stochastic matrices have maximum eigenvalue 1
- The probability π_i is the fraction of times that we visited state i as $t \to \infty$

Random walk with absorbing nodes

- Absorbing nodes: nodes from which the random walk cannot escape.

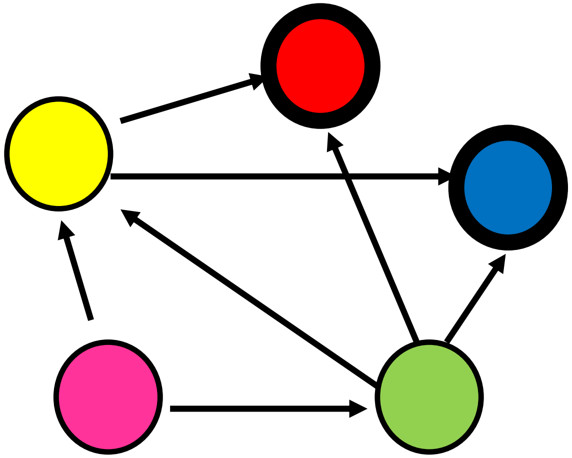


- Two absorbing nodes: the red and the blue.

P. G. Doyle, J. L. Snell. *Random Walks and Electrical Networks*. 1984

Absorption probability

- In a graph with more than one absorbing nodes a random walk that starts from a non-absorbing (transient) node t will be absorbed in one of them with some probability
  – For node t we can compute the probabilities of absorption



Absorption probabilities

- The absorption probability has several practical uses.
- Given a graph (directed or undirected) we can choose to make some nodes absorbing.
  - Simply direct all edges incident on the chosen nodes towards them and create a self-loop.
- The absorbing random walk provides a measure of proximity of transient nodes to the chosen nodes.
  - Useful for understanding proximity in graphs
  - Useful for propagation in the graph
    - E.g, on a social network some nodes are malicious, while some are certified, to which class is a transient node closer?

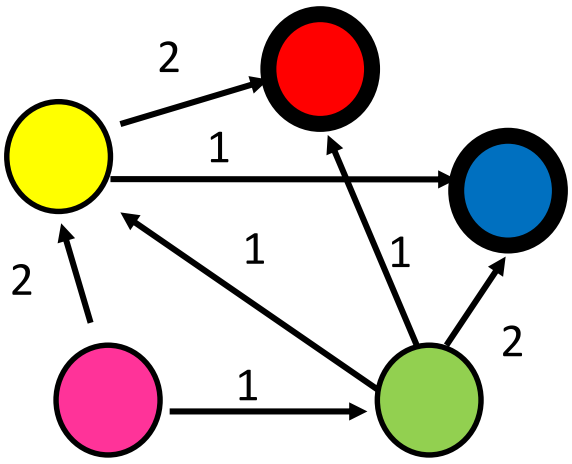
Absorption probabilities

- The absorption probability can be computed iteratively:
  - The absorbing nodes have probability 1 of being absorbed in themselves and zero of being absorbed in another node.
  - For the non-absorbing nodes, take the (weighted) average of the absorption probabilities of your neighbors
    - if one of the neighbors is the absorbing node, it has probability 1
  - Repeat until convergence (= very small change in probs)

$$P(Red|Pink) = \frac{2}{3}P(Red|Yellow) + \frac{1}{3}P(Red|Green)$$

$$P(Red|Green) = \frac{1}{4}P(Red|Yellow) + \frac{1}{4}$$

$$P(Red|Yellow) = \frac{2}{3}$$



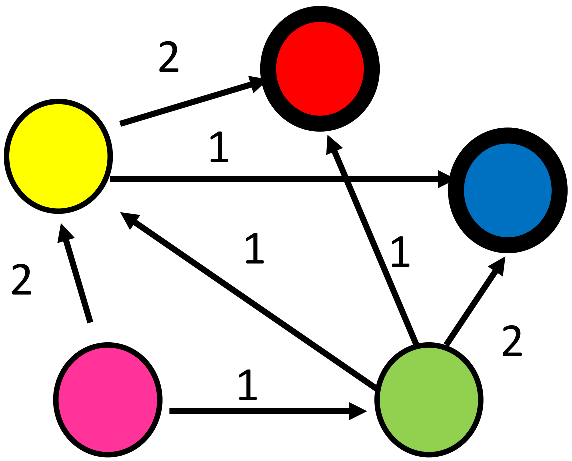
Absorption probabilities

- The absorption probability can be computed iteratively:
  - The absorbing nodes have probability 1 of being absorbed in themselves and zero of being absorbed in another node.
  - For the non-absorbing nodes, take the (weighted) average of the absorption probabilities of your neighbors
    - if one of the neighbors is the absorbing node, it has probability 1
  - Repeat until convergence (= very small change in probs)

$$P(Blue|Pink) = \frac{2}{3}P(Blue|Yellow) + \frac{1}{3}P(Blue|Green)$$

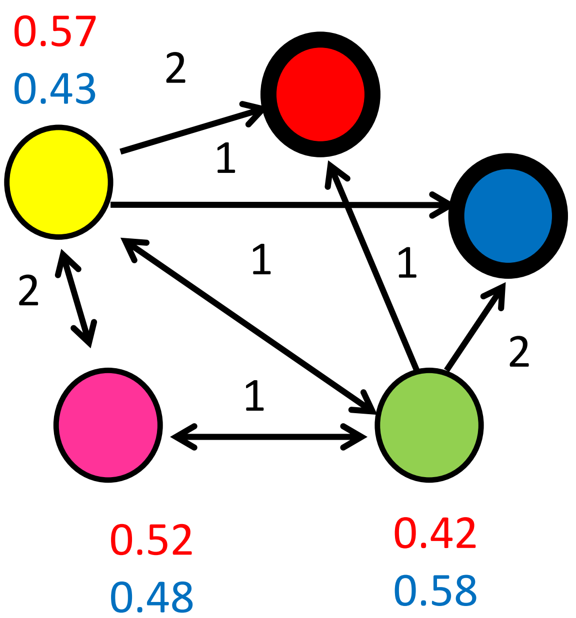
$$P(Blue|Green) = \frac{1}{4}P(Blue|Yellow) + \frac{1}{2}$$

$$P(Blue|Yellow) = \frac{1}{3}$$



Absorption probabilities

- Compute the absorption probabilities for red and blue

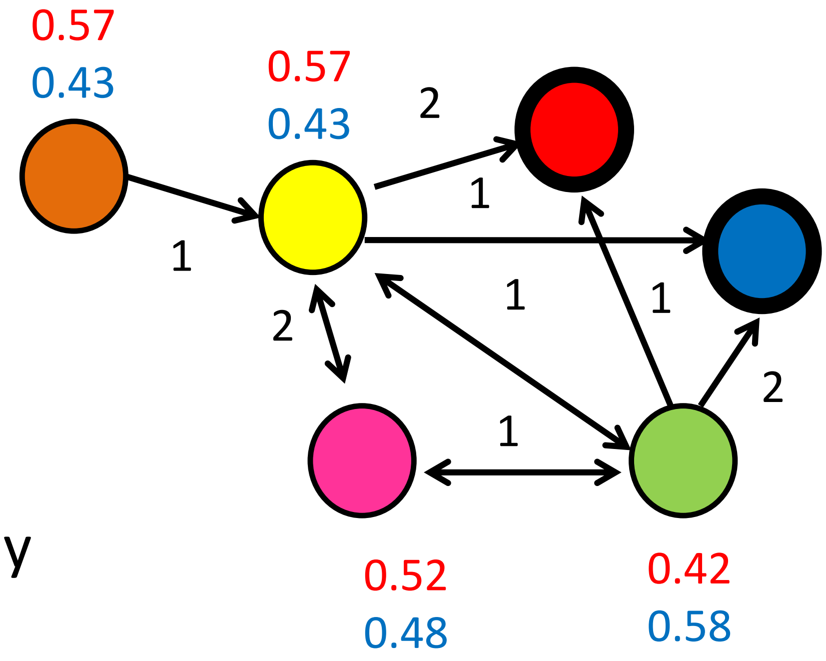


Penalizing long paths

- The orange node has the same probability of reaching red and blue as the yellow one

$$P(Red|Orange) = P(Red|Yellow)$$

$$P(Blue|Orange) = P(Blue|Yellow)$$



- Intuitively though it is further away

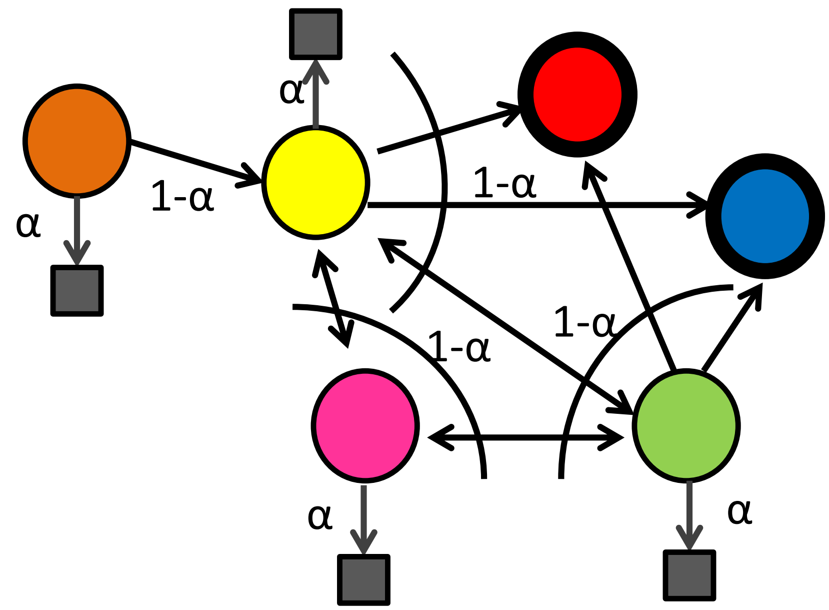
Penalizing long paths

- Add an universal absorbing node to which each node gets absorbed with probability α.

With probability α the random walk dies

With probability (1-α) the random walk continues as before

The longer the path from a node to an absorbing node the more likely the random walk dies along the way, the lower the absorbtion probability



$$P(Red|Green) = (1-\alpha)\left(\frac{1}{5}P(Red|Yellow) + \frac{1}{5}P(Red|Pink) + \frac{1}{5}\right)$$

Linear Algebra

- Our matrix looks like this

$$P = \begin{bmatrix} P_{TT} & P_{TA} \\ 0 & I \end{bmatrix}$$

- P_{TT}: transition probabilities between transient nodes
- P_{TA}: transition probabilities from transient to absorbing nodes
- Computing the absorption probabilities corresponds to iteratively multiplying matrix P with itself

Linear algebra

- The fundamental matrix

$$F = P_{TT} + P_{TT}^2 + \cdots = \sum_{i=1}^{\infty} P_{TT}^i = (1 - P_{TT})^{-1}$$

  - $F[i, j]$ = The probability of being in a transient state t_j when starting from state t_i after infinite steps
- The transient-to-absorbing matrix Q

$$Q = FP_{TU}$$

  - $Q[i, k]$ = The probability of being absorbed in absorbing state a_k when starting from transient state t_i

$$P^{\infty} = \begin{bmatrix} 0 & Q \\ 0 & I \end{bmatrix}$$

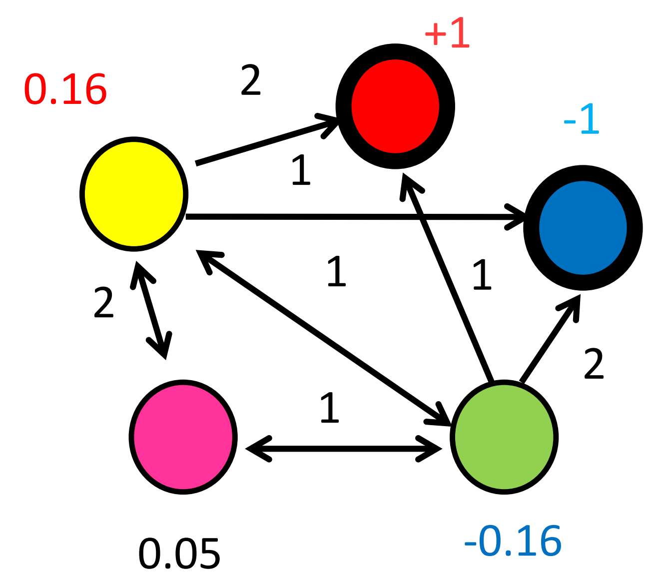
Propagating values

- Assume that Red has a positive value and Blue a negative value
- We can compute a value for all transient nodes in the same way we compute probabilities
  - This is the expected value at the absorbing node for the non-absorbing node

$$V(Pink) = \frac{2}{3}V(Yellow) + \frac{1}{3}V(Green)$$

$$V(Green) = \frac{1}{5}V(Yellow) + \frac{1}{5}V(Pink) + \frac{1}{5} - \frac{2}{5}$$

$$V(Yellow) = \frac{1}{6}V(Green) + \frac{1}{3}V(Pink) + \frac{1}{3} - \frac{1}{6}$$



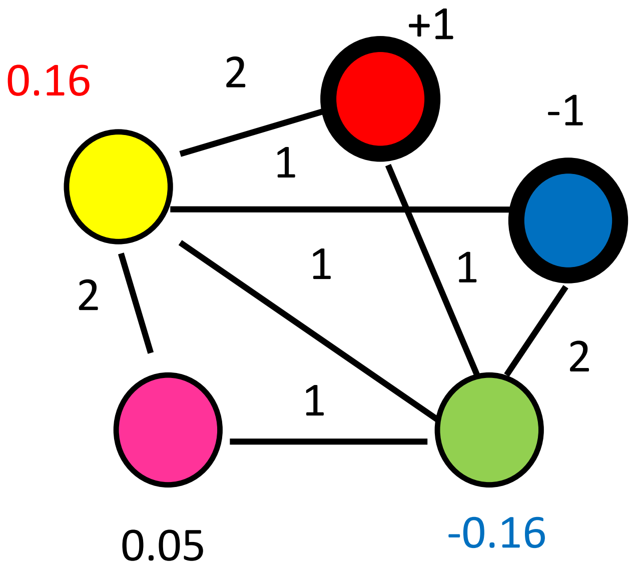
Electrical networks and random walks

- Our graph corresponds to an electrical network
- There is a positive voltage of +1 at the Red node, and a negative voltage -1 at the Blue node
- There are resistances on the edges inversely proportional to the weights (or conductance proportional to the weights)
- The computed values are the voltages at the nodes

$$V(Pink) = \frac{2}{3}V(Yellow) + \frac{1}{3}V(Green)$$

$$V(Green) = \frac{1}{5}V(Yellow) + \frac{1}{5}V(Pink) + \frac{1}{5} - \frac{2}{5}$$

$$V(Yellow) = \frac{1}{6}V(Green) + \frac{1}{3}V(Pink) + \frac{1}{3} - \frac{1}{6}$$



Linear algebra

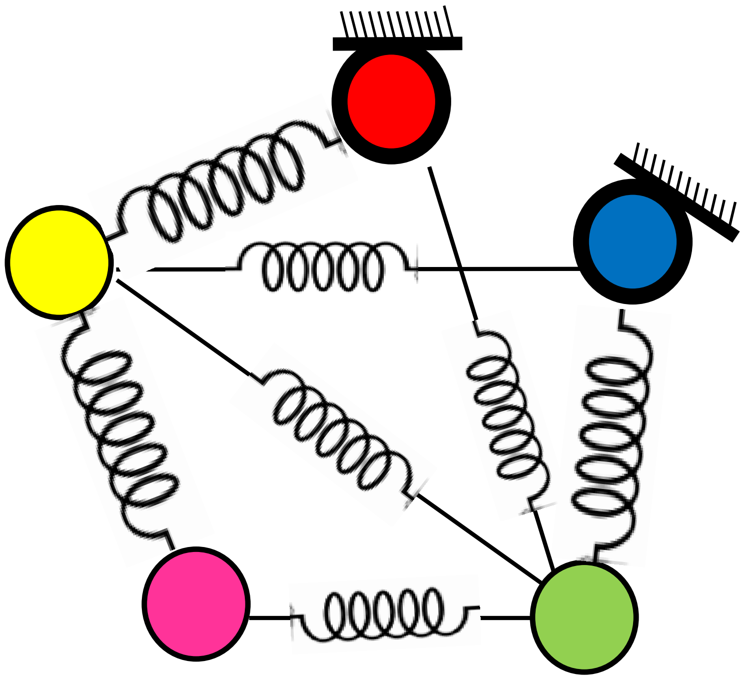
- Computation of values is essentially multiplication of the matrix Q with the vector of values of the absorbing nodes

$$\boldsymbol{v} = Q\boldsymbol{s}$$

  - $\boldsymbol{s}$: vector of values of the absorbing nodes
  - $\boldsymbol{v}$: vector of values of the transient nodes

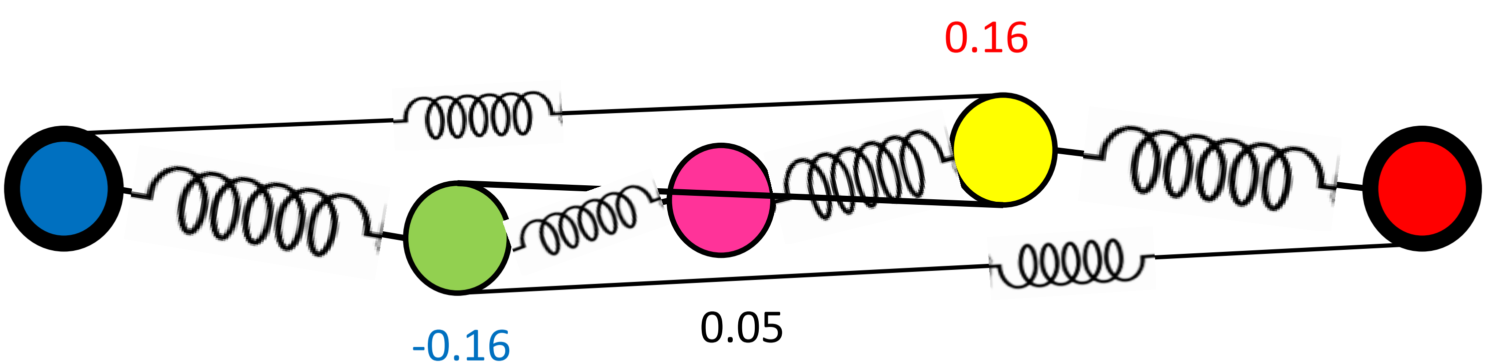
Springs and random walks

- Our graph corresponds to an spring system
- The Red node is pinned at position +1, while the Blue node is pinned at position -1 on a line.
- There are springs on the edges with hardness proportional to the weights
- The computed values are the positions of the nodes on the line



Springs and random walks

- Our graph corresponds to an spring system
- The Red node is pinned at position +1, while the Blue node is pinned at position -1 on a line.
- There are springs on the edges with hardness proportional to the weights
- The computed values are the positions of the nodes on the line

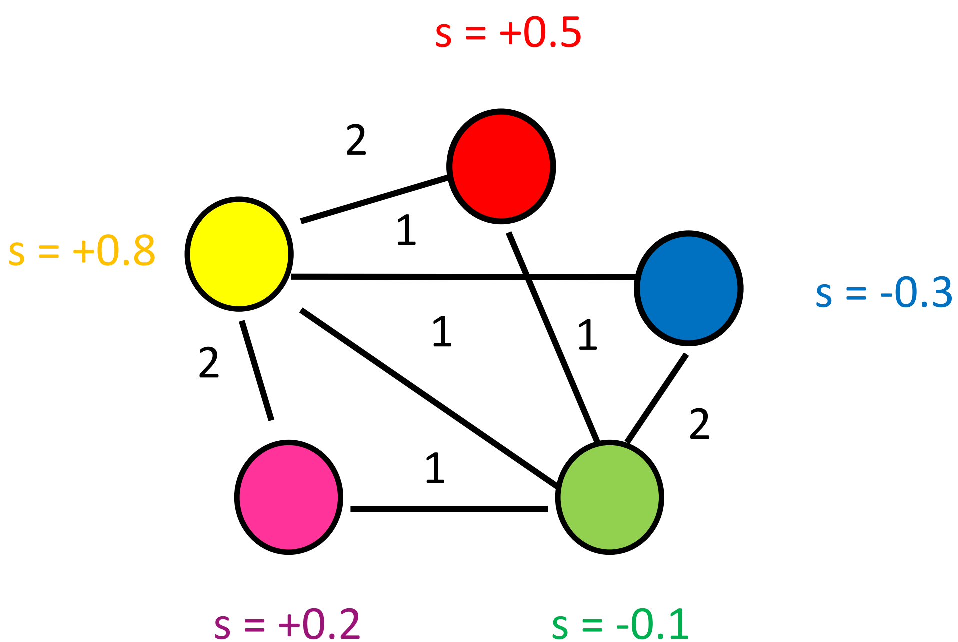


Application: Transductive learning

- If we have a graph of relationships and some labels on some nodes we can propagate them to the remaining nodes
  - Make the labeled nodes to be absorbing and compute the probability for the rest of the graph
  - E.g., a social network where some people are tagged as spammers
  - E.g., the movie-actor graph where some movies are tagged as action or comedy.
- This is a form of semi-supervised learning
  - We make use of the unlabeled data, and the relationships
- It is also called transductive learning because it does not produce a model, but just labels the unlabeled data that is at hand.
  - Contrast to inductive learning that learns a model and can label any new example

Back to opinion formation

- The value propagation we described is closely related to the opinion formation process/game we defined.
  – Can you see how we can use absorbing random walks to model the opinion formation for the network below?



Reminder:

$$z_i = \frac{s_i + \sum_{j \in N(i)} w_{ij} z_j}{1 + \sum_{j \in N(i)} w_{ij}}$$

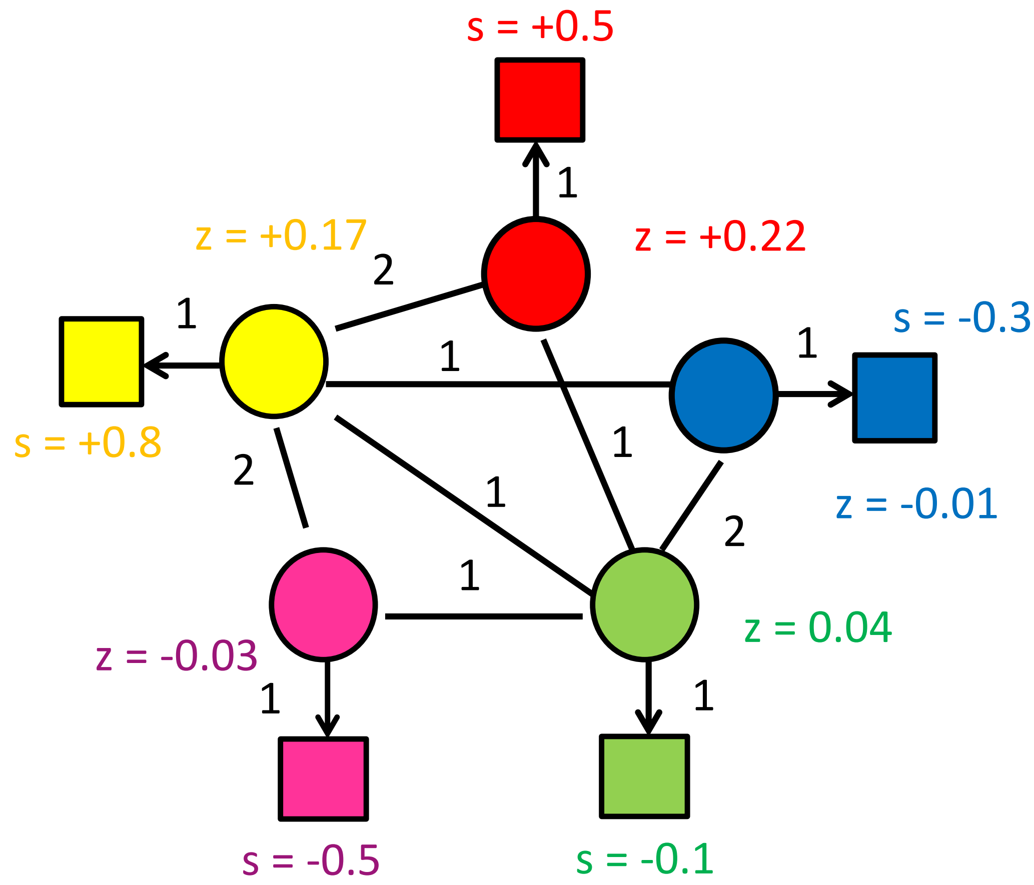
Opinion formation and absorbing random walks

One absorbing node per user with value the intrinsic opinion of the user

One transient node per user that links to her absorbing node and the transient nodes of her neighbors

The expressed opinion for each node is computed using the value propagation we described

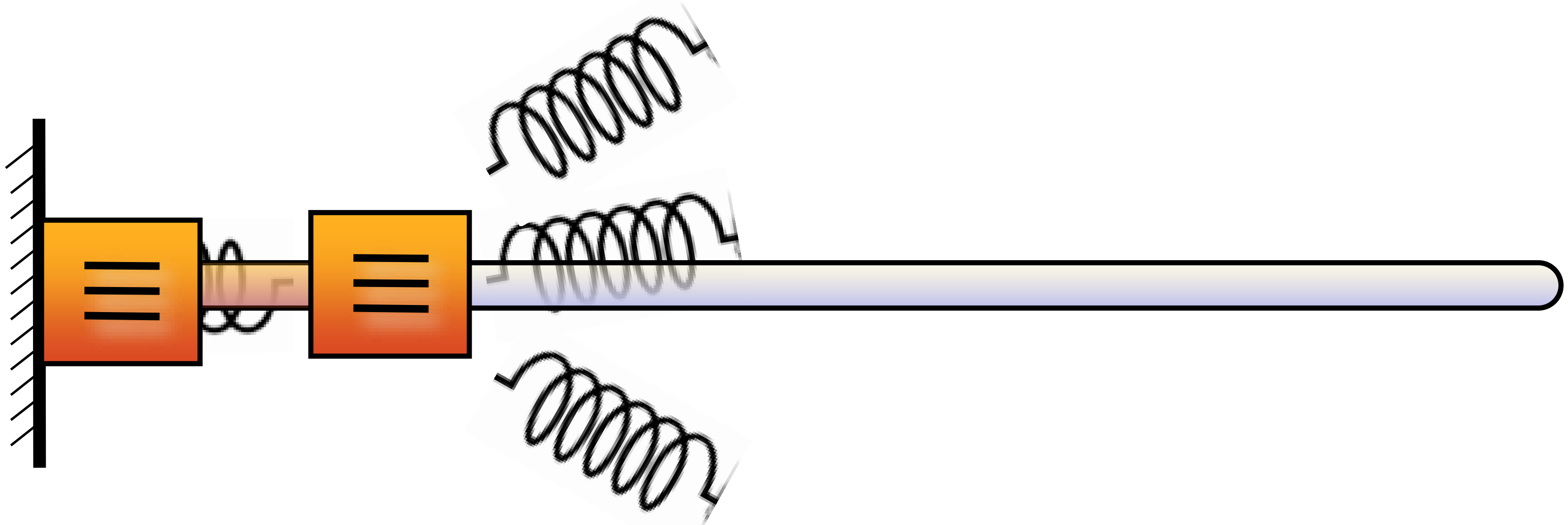
- Repeated averaging

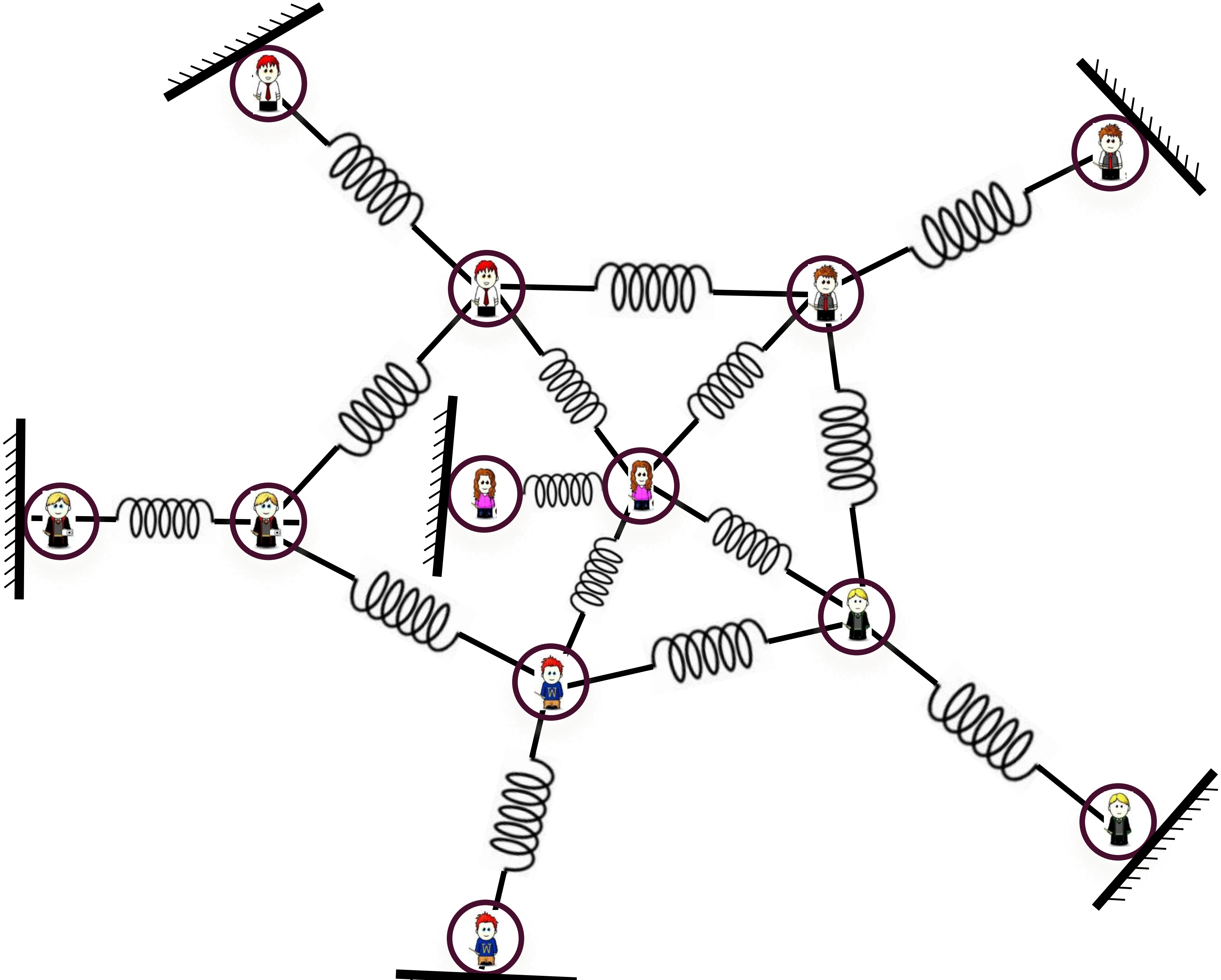


It is equal to the expected intrinsic opinion at the place of absorption

Opinion of a user

- For an individual user u
  - u's absorbing node is a stationary point
  - u's transient node is connected to the absorbing node with a spring.
  - The neighbors of u pull with their own springs.





Opinion maximization problem

- Public opinion:

$$g(z) = \sum_{i \in V} z_i$$

- Problem: Given a graph G, the given opinion formation model, the intrinsic opinions of the users, and a budget k, perform k interventions such that the public opinion is maximized.

- Useful for image control campaign.

- What kind of interventions should we do?

Possible interventions

1. Fix the expressed opinion of k nodes to the maximum value 1.
   - Essentially, make these nodes absorbing, and give them value 1.

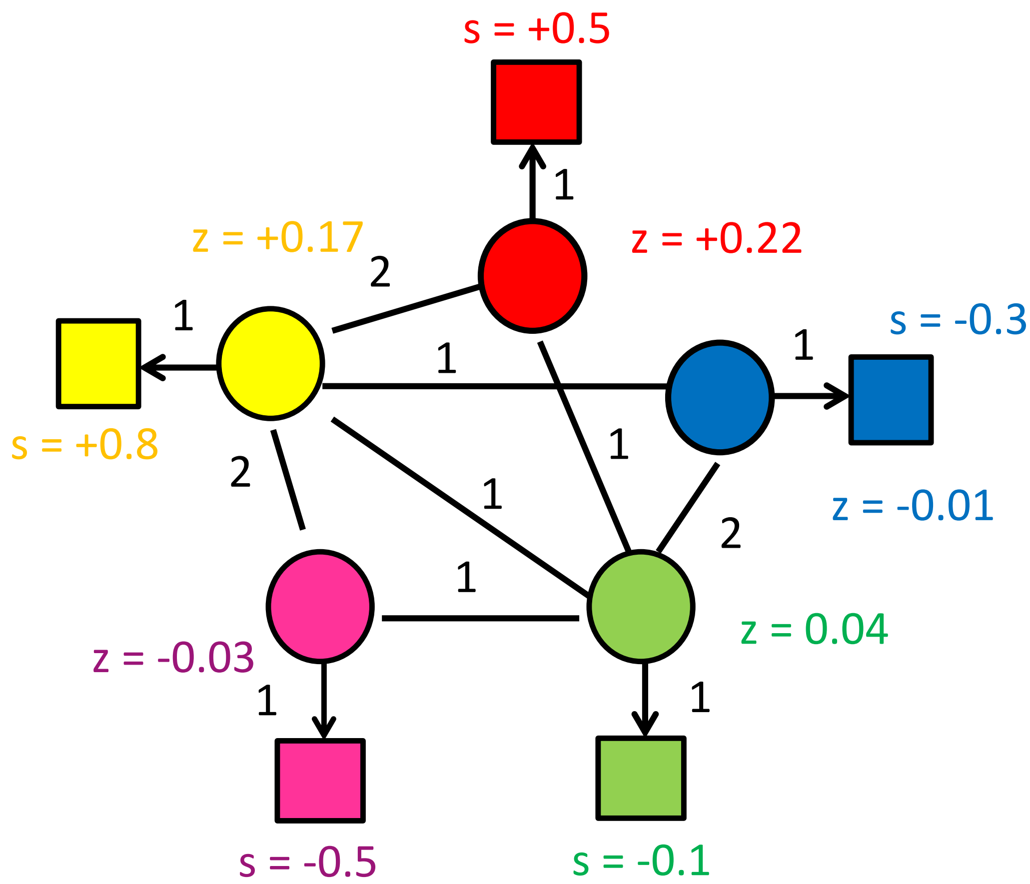
2. Fix the intrinsic opinion of k nodes to the maximum value 1.
   - Easy to solve, we know exactly the contribution of each node to the overall public opinion.

3. Change the underlying network to facilitate the propagation of positive opinions.
   - For undirected graphs this is not possible

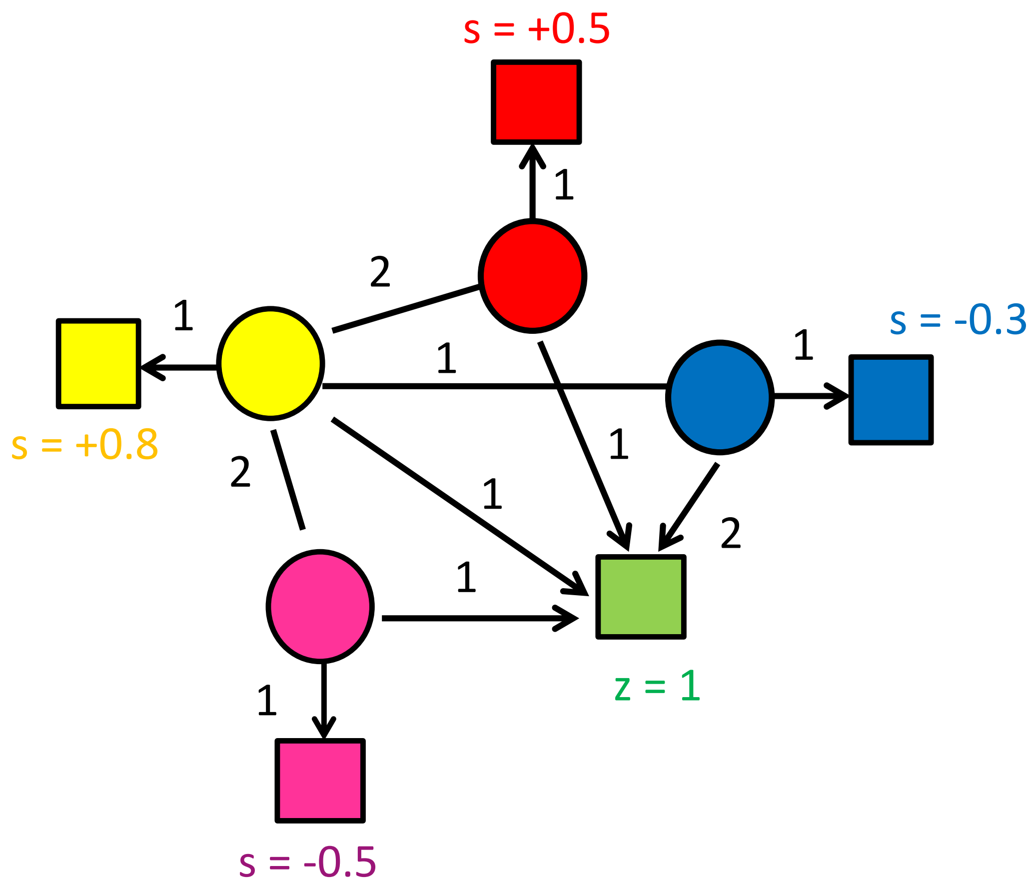
$$g(z) = \sum_i z_i = \sum_i s_i$$

   - The overall public opinion does not depend on the graph structure!
   - What does this mean for the wisdom of crowds?

Fixing the expressed opinion



Fixing the expressed opinion



Opinion maximization problem

- The opinion maximization problem is NP-hard.
- The public opinion function is monotone and submodular
  - The Greedy algorithm gives an $\left(1 - \frac{1}{e}\right)$-approximate solution
- In practice Greedy is slow. Heuristics that use random walks perform well.

A. Gionis, E. Terzi, P. Tsaparas. *Opinion Maximization in Social Networks*. SDM 2013

Additional models

- Ising model
- Voter model
- Bounded confidence models
- Axelrod cultural dynamics model

A Physics-based model

- The Ising ferromagnet model:
  - A user i is a "spin" s_i that can assume two values: ±1
  - The total energy of the system is

$$H = -\frac{1}{2}\sum_{\langle i,j\rangle} s_i s_j$$

  Defined over the neighboring pairs
  - A spin is flipped with probability $\exp(-\frac{\Delta E}{T})$ where ΔE is the change in energy, and T is the "temperature" of the system.
- The model assumes no topology
  - Complete graph (all-with-all), or regular lattice.
- For low temperatures, the system converges to a single opinion

The Voter model

- Each user has an opinion that is an integer value (usually opinions are in {0,1} but multiple opinion values are also possible).
- Opinion formation process:
  - At each step we select a user at random
  - The user selects one of its neighbors at random (including herself) and adopts their opinion
- The model can be proven to converge for certain topologies.

Bounded confidence model

- Confirmation bias: People tend to accept opinions that agree with them
  - "Why facts don't change our minds" (New Yorker)

- Bounded Confidence model: A user i is influenced by a neighbor j only if

$$\left|z_i - z_j\right| \leq \epsilon$$

for some parameter ϵ

Bounded Confidence models

- Defuant model: Given a parameter μ at time t, a randomly selected user i selects a neighbor j at random, and if $\left|z_i^t - z_j^t\right| \leq \epsilon$ their opinions are updated as:

$$z_i^{t+1} = z_i^t + \mu(z_j^t - z_i^t)$$
$$z_j^{t+1} = z_j^t + \mu(z_i^t - z_j^t)$$

Similar to Voter model

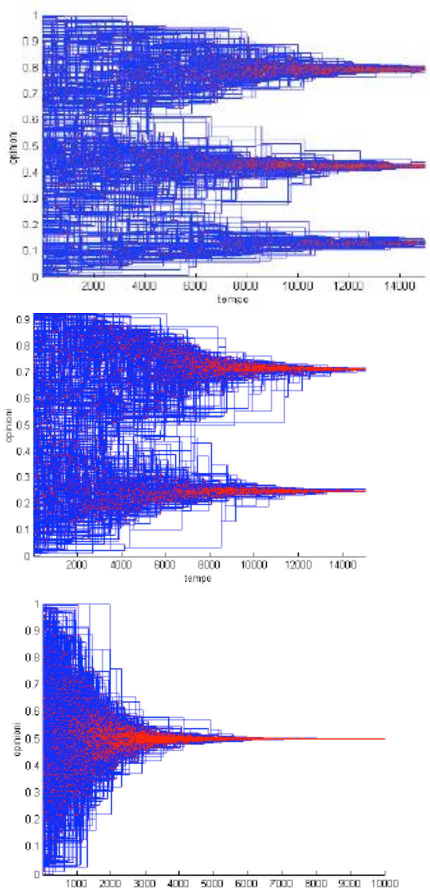
- Hegselmann-Krause (HK) model: Each node i updates their opinions as the average of the opinions of the neighbors that agree with them

$$z_i^t = \frac{w_{ii} z_i^{t-1} + \sum_{j \in N(i): \left|z_i^t - z_j^t\right| \leq \epsilon} w_{ij} z_j^{t-1}}{w_{ii} + \sum_{j \in N(i): \left|z_i^t - z_j^t\right| \leq \epsilon} w_{ij}}$$

Similar to DeGroot model

Bounded Confidence models

- Depending on the parameter ϵ and the initial opinions, bounded confidence models can lead to plurality (multiple opinions), polarization (two competing opinions), or consensus (single opinion)



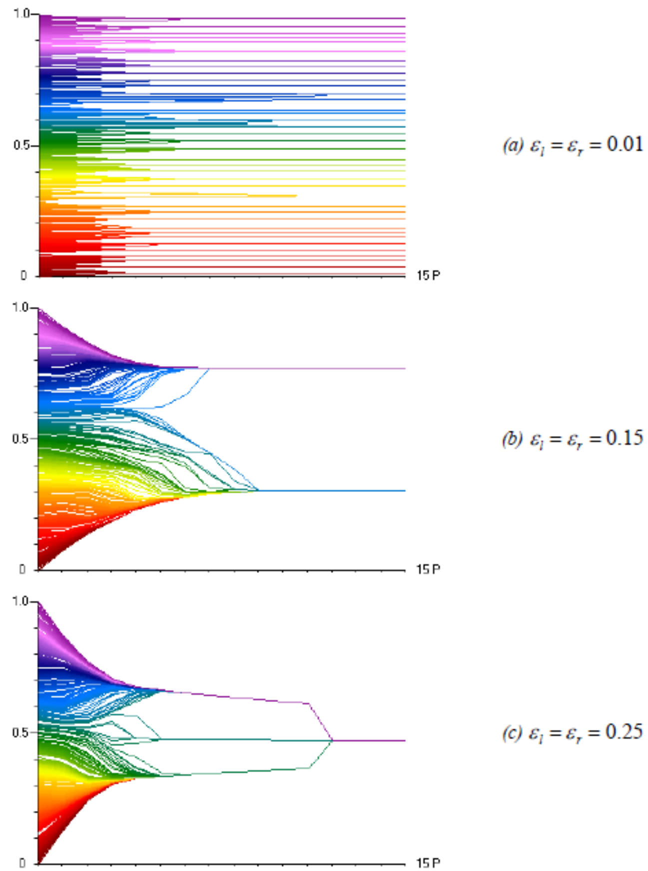


Figure 2: Stops while walking along the diagonal

Axelrod model

- Cultural dynamics: Goes beyond single opinions, and looks at different features/habits/traits
  - Tries to model the effects of social influence and homophily.
- Model:
  - Each user i has a vector σ_i of F features
  - A user i decides to interact with user j with probability

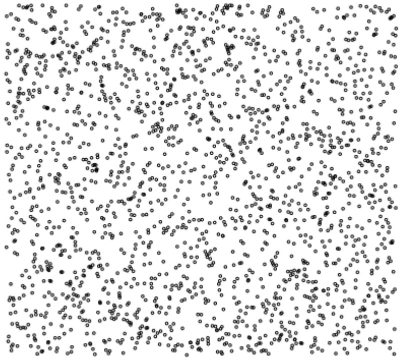
$$\omega_{ij} = \frac{1}{\mathrm{F}} \sum_{f=1}^{F} \delta(\sigma_i(f), \sigma_j(f))$$

Fraction of common features

  - If there is interaction, the user changes one of the disagreeing features to the value of the neighbor
- The state where all users have the same features is an equilibrium, but it is not always reached (cultural pockets)

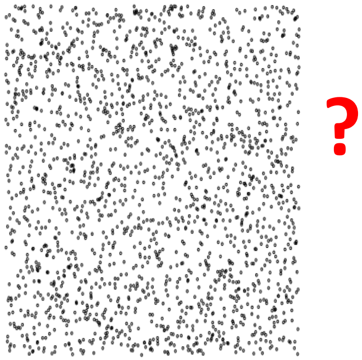
Empirical measurements

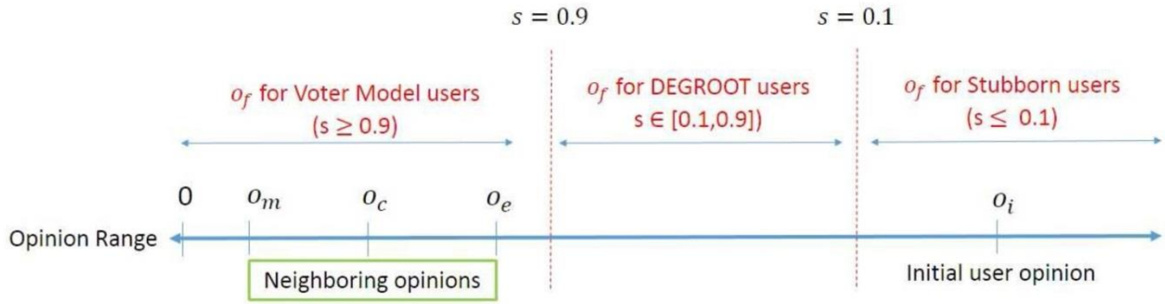
- There have been various experiments for validating the different models in practice



- Das, Gollapudi, Munagala (WSDM 2014)
  - User surveys:
    - estimate number of dots in images
    - Estimate annual sales of various brands.
  - For each survey:
    - Users asked to provide initial answers on all questions in the survey
    - Then, each user shown varying number of (synthetic) neighboring answers.
    - Users given opportunity to update their answers

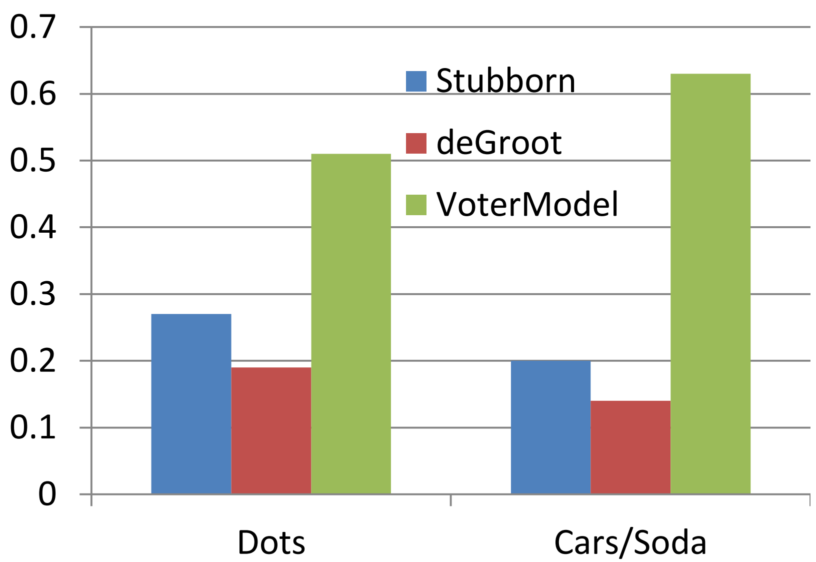
Online User Studies



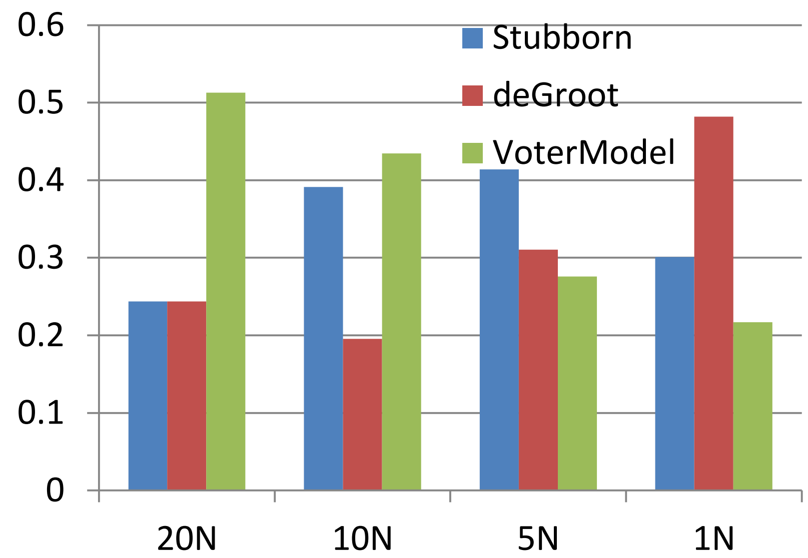


- Define $s = \frac{|o_i - o_f|}{|o_i - o_e|}$
  - (o_i: original opinion, o_f: final opinion, o_e: closest neighboring opinion)
- User behavior categorized as:
  - Stubborn ($s < 0.1$)
  - DeGroot ($0.1 < s < 0.9$)
  - Voter ($s > 0.9$)

Voter vs DeGroot



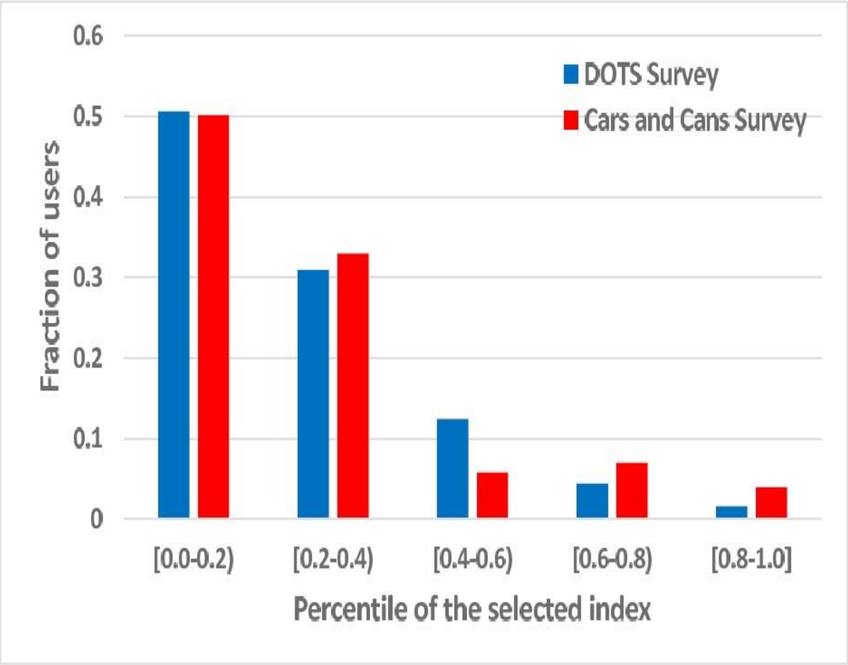
Distribution over stubborn, deGroot and voter



Effect of number of neighboring opinions

- Voter model is prevalent for large number of neighbors,
- DeGroot becomes more prevalent for smaller number of neighbors

Biased Conforming Behavior



- Adoption of neighboring opinions not uniform random (unlike Voter Model)
- Users give higher weights to “close by” opinions

Other problems related to opinion formation

- Modeling polarization
  – Understand why extreme opinions are formed and people cluster around them
- Modeling herding/flocking
  – Understand under what conditions people tend to follow the crowd
- Computational Sociology
  – Use big data for modeling human social behavior.

R. Hegselmann, U. Krause. *Opinion Dynamics and Bounded Confidence. Models, Analysis, and Simulation*. Journal of Artificial Societies and Social Simulation (JASSS) vol.5, no. 3, 2002

Acknowledgements

- Many thanks to Evimaria Terzi, Aris Gionis and Sreenivas Gollapudi for their generous slide contributions.

References

- M. H. DeGroot. *Reaching a consensus*. J. American Statistical Association, 69:118–121, 1974.
- N. E. Friedkin and E. C. Johnsen. *Social influence and opinions*. J. Mathematical Sociology, 15(3-4):193–205, 1990.
- D. Bindel, J. Kleinberg, S. Oren. *How Bad is Forming Your Own Opinion?* Proc. 52nd IEEE Symposium on Foundations of Computer Science, 2011.
- P. G. Doyle, J. L. Snell. *Random Walks and Electrical Networks*. 1984
- A. Gionis, E. Terzi, P. Tsaparas. *Opinion Maximization in Social Networks*. SDM 2013
- R. Hegselmann, U. Krause. *Opinion Dynamics and Bounded Confidence. Models, Analysis, and Simulation*. Journal of Artificial Societies and Social Simulation (JASSS) vol.5, no. 3, 2002
- C. Castellano, S. Fortunato, V. Loreto. *Statistical Physics of Social Dynamics*, Reviews of Modern Physics 81, 591-646 (2009)
- A. Das, S. Gollapudi, K. Munagala, *Modeling opinion dynamics in social networks*. WSDM 2014