Social and Information Networks

University of Toronto CSC303
Winter/Spring 2022

Week 7: Feb 28 - March 4 (2022)

This week's agenda

- Twitter rumour cascades
- Structural virality
- Threshold model
  - Complete cascades
  - Blocking clusters
  - A first look at selecting initial adopters

Chapter 19: Influence spread in a social network

- We begin a study of the spread/diffusion of products/influence in an existing social network (Chapter 19). This is in contrast to the population wide influence spread that we are passing over in Chapters 16 and 17. Chapter 18 (on power laws) also dealt with population wide influence phenomena.

The chapters preceding chapter 19

- In Chapters 16 (information cascades), 17 (direct benefit effects), and 18 (rich get richer models) there isn't a social network per se.
- These chapters dealt with population wide effects. Although :
  - One can construe Chapter 16 as taking place in a network where the i^{th} individual is connected to all $i-1$ previous individuals.
  - Chapter 17 can be construed as taking place in the complete graph network. Information about the entire population impacts decisions.
  - In Chapter 18 we studied studied a random process (e.g., link creation) by which an information network can grow. We also studied in this chapter, an example (music downloading in the Salganik et al experiment) where we can identify how the presence of population wide information will influence an outcome. Like Chapter 16, we can think of this as taking place in a social network where the i^{th} person knows some global information about the preceding $i-1$ individuals.
- But basically these are population wide effects absent from an existing social network where influence spreads without any global information.

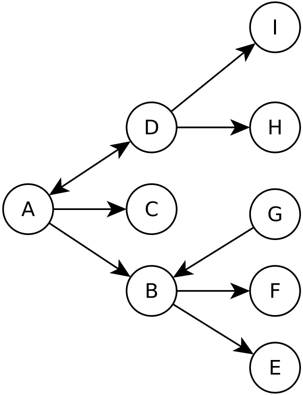
Aside: It is interesting to contrast the herding effect in chapter 16 with the impact of influence in the Salganik et al experiment in Chapter 18.

Social network effects

- Now we wish to consider an existing social network where edges (ties) between individuals represent some sort of friendship/relationship.
- This takes us back to concepts introduced in Chapters 3 and 4.
- There we saw the contrast between
  - selection (we tend to be friends with people of similar backgrounds, geography, interests)
  - social influence (we join clubs, are influenced) by our friends/relations.
- Rather than link creation (e.g., selection), we will now study spread (e.g., influence)
- The goal (as throughout the course) is to qualitatively understand a process or observed phenomena in a highly stylized (but hopefully still interesting) setting.
- We will (as usual) be interested in what kind of general conclusions can be inferred from such an understanding.

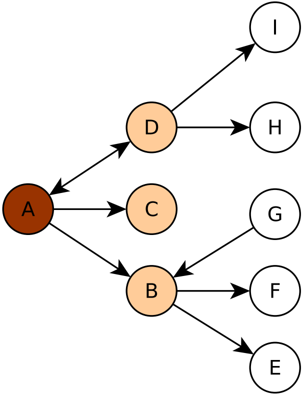
Example: Spread of news through Twitter

- Consider a graph where nodes are Twitter accounts
  - The directed edge $\langle A, B \rangle$ indicates that B follows A
- News "spreads" through this network via re-tweets.
  - Let light orange nodes those who hear the news
  - Let dark orange nodes be those who repeat the news



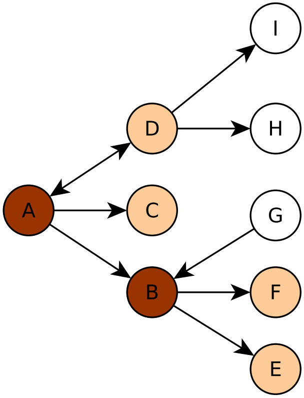
Example: Spread of news through Twitter

- Consider a graph where nodes are Twitter accounts
  - The directed edge $\langle A, B \rangle$ indicates that B follows A
- News "spreads" through this network via re-tweets.
  - Let light orange nodes those who hear the news
  - Let dark orange nodes be those who repeat the news



Example: Spread of news through Twitter

- Consider a graph where nodes are Twitter accounts
  - The directed edge $\langle A, B \rangle$ indicates that B follows A
- News "spreads" through this network via re-tweets.
  - Let light orange nodes those who hear the news
  - Let dark orange nodes be those who repeat the news



Example: Spread of news through Twitter

- How does this transmission behave? Are there differences between the transmission of true & false information?

  "*Falsehood will fly, as it were, on the wings of the wind, and carry its tales to every corner of the earth; whilst truth lags behind; her steps, though sure, are slow and solemn, and she has neither vigour nor activity enough to pursue and overtake her enemy*"
  – Thomas Francklin

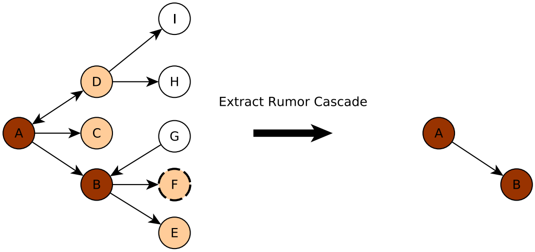
  "*A lie can run round the world before the truth has got its boots on*"
  – Sir Terry Pratchett, The Truth

Example: Spread of news through Twitter

- We'll be looking at an interesting paper by Vosoughi et al. looking at the spread of real and fake news through Twitter (`https://science.sciencemag.org/content/sci/359/6380/1146.full.pdf`)
- They defined news to be true (resp. false) if it was verified (resp. rejected) by one of six independent fact checking organizations
  - snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, urbanlegends.about.com

Example: Spread of news through Twitter

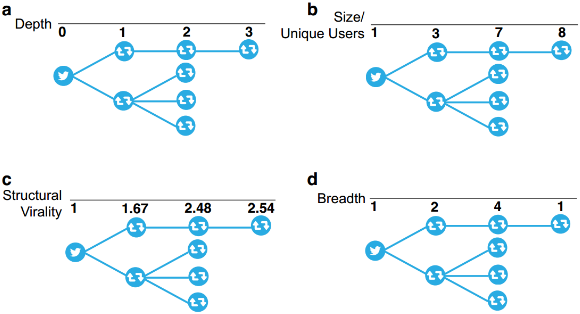
- They extracted *Rumour Cascades* on Twitter (i.e. the subgraph of the original poster & the retweeting nodes)
  - Looked for some tweet F, that replied to some tweet B with a fact-checking link. The originating tweet A was identified (either B, or the original post that B retweeted) and then cascade was extracted by finding A and A's retweets
  - A combination of NLP and manual inspection was used to check that the original tweet actually related to the fact check link



- Their final dataset contained 126,000 stories, tweeted by 3 million people more than 4.5 million times.

Measuring a Rumour Cascade

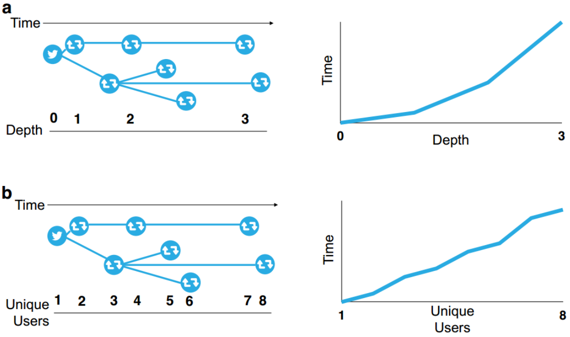
- Question: For a given true or false rumour we can have multiple cascades. What may we want to measure in a given cascade? How could this be interesting or of value?
- Vosoughi et al. looked at the static measures of depth, size, maximum breadth (the number of users at a given depth), and *structural virality*



[From Vosoughi et al.]

Measuring a Rumour Cascade

- Question: For a given true or false rumour we can have multiple cascades. What may we want to measure in a given cascade? How could this be interesting or of value?
- Vosoughi et al. looked at the dynamic measures of depth over time, users over time, breadth vs. depth, and size vs. depth



[From Vosoughi et al.]

Structural Virality

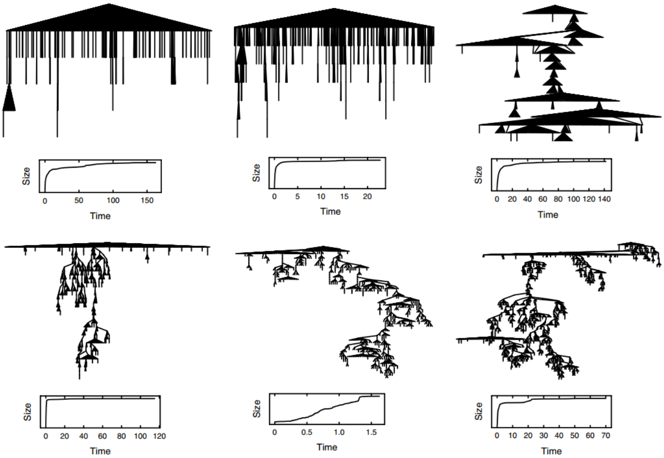
- Structural virality is a measure meant to distinguish transmission via broadcast (large burst or bursts) or rapid peer-to-peer spreading (exponential growth over time)

Figure 1 **A Schematic Depiction of Broadcast vs. Viral Diffusion, Where Nodes Represent Individual Adoptions and Edges Indicate Who Adopted from Whom**

[From The Structural Virality of Online Diffusion, Goel et al., 2014]

Structural Virality

Figure 3 A Random Sample of Cascades Stratified and Ordered by Increasing Structural Virality, Ranging from 2 to 50



[From The Structural Virality of Online Diffusion, Goel et al., 2014]

How to Measure Structural Virality?

- Some desirable properties of a measurement of structural virality:
  - For a fixed size, structural virality should increase with the average branching factor
  - For a fixed average branching factor, structural virality should increase with depth
  - All pure broadcast structures should be equally viral, regardless of their size
- Question: How can we try to measure structural virality? Can we satisfy all these desiderata?
- Depth satisfies the last two, but violates the first (e.g. long chains)
- Average depth solves chains, but is still problematic

Structural Virality

Structural Virality

- Some desirable properties of structural virality:
  - For a fixed size, structural virality should increase with the average branching factor
  - For a fixed average branching factor, structural virality should increase with depth
  - All pure broadcast structures should be equally viral, regardless of their size
- Structural virality is defined as the average distance between all pairs of nodes

$$\text{virality}(G) = \frac{1}{|V|(|V|-1)} \sum_{u,v \in V} d(u,v)$$

- Equivalently, it can be viewed as the average distance to a node, averaged over all nodes
- It still has pathological cases, but it is empirically useful

Example: Spread of news through Twitter

- Looking at our static measures, we can see that false news cascades travel deeper, reach more people, and have greater breadth and structural virality
- Looking at our dynamic measures, we can see that, on average, fake news grows in depth and size faster than real news

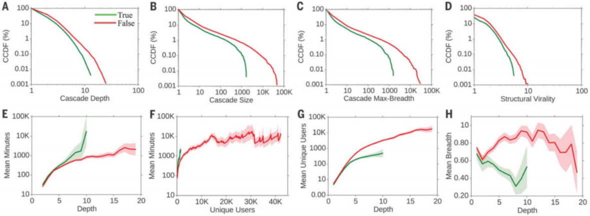


Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades. (**A**) Depth. (**B**) Size. (**C**) Maximum breadth. (**D**) Structural virality. (**E** and **F**) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (**G**) The number of unique Twitter users reached at every depth and (**H**) the mean breadth of true and false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level (i.e., cascades belonging to the same rumor were clustered together; see supplementary materials for additional details).

- All of these features were significantly different, and seem to indicate that Truth does indeed tarry with it's boots
- Question: What could be possible causes of this?

Fake vs. Real news spread: Structural Differences?

- Differences in the underlying following-followee network around nodes prone to spreading falsehood could explain the result
- Could these nodes have higher out-degree, or in-degree?
- Vosoughi et al. found that the nodes spreading false information tended to have significantly *fewer* followers and followees
- Furthermore, at the individual level they were significantly less active, and on Twitter for significantly less time

Fake vs. Real news spread: Blame the bots?

- Bots in the network could be encouraging spread
- In effect, these nodes would be selectively more contagious
- The authors removed bots from their data and recalculated
- They found that bots increase spread, however both real and fake news were amplified equally

Fake vs. Real news spread

- The authors did an analysis of the text of real and fake news, and found that fake news was consistently more novel under various metrics
- The authors concluded that rather than structural factors or individual characteristics, the greater spread of misinformation comes from individuals being more likely to transmit it
- In short, a lie *really can* go around the world before the truth has got it's boots on
- More recent work by Meyers et al. (`https://doi.org/10.1007/978-3-030-61841-4_10`) has attempted to exploit these structural differences in rumour cascades to identify fake news
- Furthermore, Meyers et al. found that although the individual cascades may be smaller, true stories tend to have a higher number of cascades, resulting in truth reaching more people overall, and resulting in truth remaining in circulation for longer on Twitter
- So indeed, at least on Twitter, it appears that the steps of truth are sure, if slow and solemn

Models of influence spread/diffusion

- One of the most important themes of the text (and CSC303) is that we construct models to gain insight.
  - Our models are often (maybe always) very simplified given the complexity of real social and economic networks.
  - There is always a tradeoff between the adherence to reality and our ability to analyse and gain insight.
- How we model diffusion in a social network will clearly depend on what product, idea, membership, etc. we are considering.
- There are many assumptions as to how products, ideas, influence are spread in a social network and what are the set of individual alternatives.
- The main emphasis in Chapter 19 is on a very simple process of diffusion where each person has 2 alternative decisions:
  1. stay with a current "product" B
  2. or switch to a (new) product A.

A simple model of diffusion in a social network

- Let's assume that we are making decisions based on the direct benefit of being coordinated with our friends beyond any intinsic value associated with the decision (e.g. when the decision is the purchase of an item).

- A standard example is what messenger application we might choose to use to the extent that we are mostly influenced by our friends rather than by general population wide usage (e.g., do you use SMS, Skype, Zoom, Teamviewer, Jitsi Meet, Microsoft Teams, email, Pidgin, Slack, Discord, Element, NeoChat, Quercus messages, Snapchat, Steam, Telegram, Instagram Direct, WhatsApp, iMessage, Facebook Messenger, WeChat, Signal, AOL instant messenger, Google Hangouts . . .). What influences you most? Friends or general population information?
  - Choosing between two weekly television shows that occur at the same time or who to vote for are other examples.

A simple model of diffusion in a social network

- In fact, the model given in this chapter dictates that certain decisions (i.e. to change from B to A) are irreversible.
  - The text calls this a "progressive process" in the sense that it progresses in only one direction. Any good examples of truly (or *essentially*) irreversible decisions?
  - For example, the decision to get a tattoo.

A threshold model for spread

- We assume that some number of individuals are enticed (at some time $t = 0$) to adopt a new product A.
- Outside of these "initial adopters", we assume all other individuals in the network are initially using a different product B (or equivalently this is the first product in a given market).
- This is not really a competitive influence model as B is not really competing. (More comments later.)
- The first model we consider for diffusion is that every node v has a threshold q (in absolute or relative terms) for how many of its neighbors must have adopted product A before v adopts A.

Threshold model (continued)

- For simplicity the text initially assumes that every node v (i.e. individual) in the network has the same threshold but then later explains how to deal with individual thresholds.
- If at some time t, the threshold for a node v has been achieved, then by time time $t + 1$, v will adopt product A.
- If the threshold has not been reached then v decides not to adopt A at this time.

Note

Although it is not explicitly stated, the initial adopters never reverse their adoption.

- Given these model assumptions, adopting A is irreversible for all nodes in the network.

Determining a (relative) threshold

- One way (some might say is usually the best way) to reason about a plausible threshold for a node is to view one's decision in economic terms.

- Specifically for every edge (v, w) in the network suppose
  - There is payoff a to v and w if both v and w have adopted product A.
  - There is payoff b to v and w if both v and w have adopted product B.
  - A zero payoff when v and w do not currently utilize the same product.

- This determines a simple coordination game.

$$
\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{w} \\
 & & A & B \\
\hline
v & A & a, a & 0, 0 \\
\hline
 & B & 0, 0 & b, b \\
\hline
\end{array}
$$

Figure: $A - B$ coordination [Fig 19.1, E&K]

Coordination game induces threshold

- Suppose node v has not yet adopted A at time t, but a fraction p of the $d(v)$ neighbors of v have already adopted A, then:
  - By switching, the payoff to v is $p \times d(v) \times a$.
  - By staying with B, v has payoff $(1 - p) \times d(v) \times b$.

- Thus node v will switch to A if

$$p \times d(v) \times a \geq (1 - p) \times d(v) \times b$$

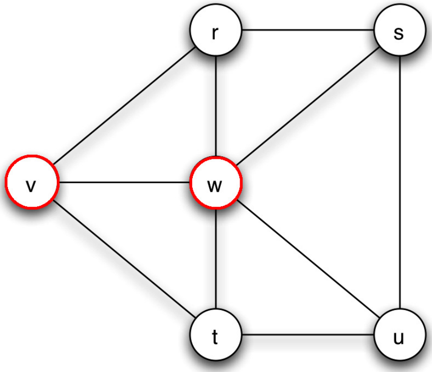
  (for simplicity say v switches when payoffs are equal).

- This is then equivalent to saying that v will switch whenever p is at least $\frac{b}{a+b} = q$ which is then the relative threshold.

- That is, whenever there is at least a (threshold) fraction q of the neighbours of node v that have adopted A, then v will also adopt A.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

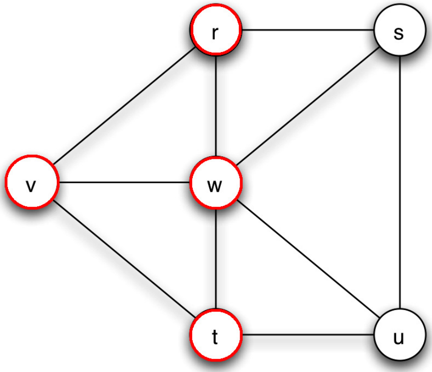


$t = 0$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

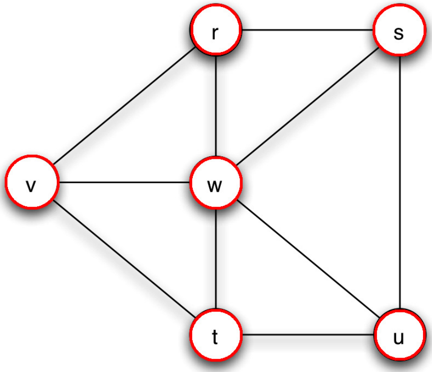


$t = 1$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

The process unfolds (example: $a = 3$ and $b = 2$)

[Fig 19.3, E&K]

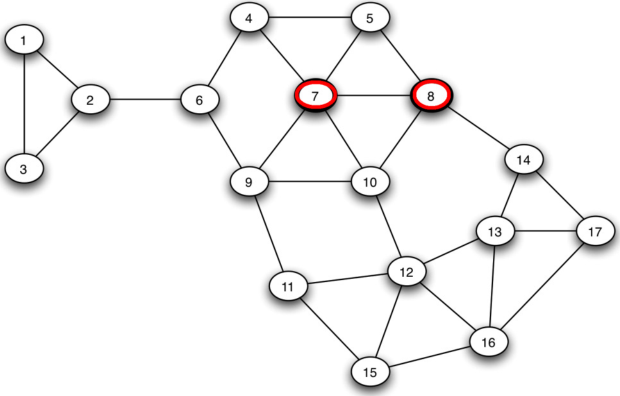


$t = 2$

- A node adopts A if and only if the threshold $q = \frac{b}{a+b} = 2/5$ is reached.
- Two nodes v and w are initial adopters.

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

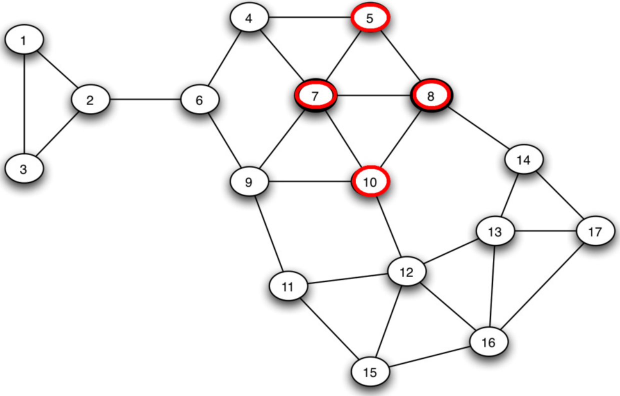


$t = 0$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

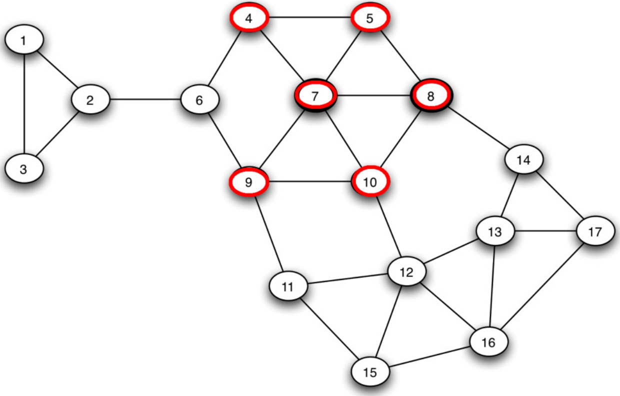


$t = 1$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.

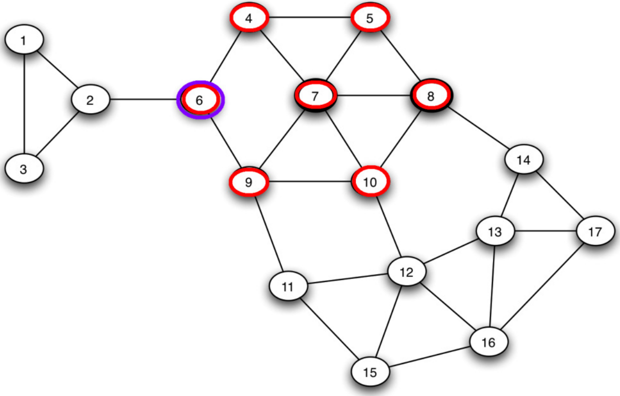


$t = 2$

[Fig 19.4, E&K]

Complete cascades vs tightly-knit communities (example: $a = 3$, $b = 2$, $q = 2/5$)

- The previous example showed a complete cascade where all nodes eventually adopt A.
- In the next example, "tightly-knit communities" block the spread.



$t = 3$

[Fig 19.4, E&K]

Factors determining the rate and extent of diffusion in a social network

1. The structure of the network.

2. The relative payoffs vs costs for adopting a new product.
   - We haven't spoken of costs yet but we usually do have a cost for adopting a new product.
   - We can introduce such a cost into the model by saying that v will not adopt the new A unless

$$p \times d(v) \times a - \text{cost} \geq (1-p) \times d(v) \times b$$

   - We could also add intrinsic values for A and B to both sides of the above inequality to determine the threshold for v adopting A.

3. The choice of initial adopters.
   - This raises an interesting computational question as to how to select the most influential nodes (within some budgetary constraint).

Defining a tightly-knit community

- We want to show that not only do tightly-knit communities cause a cascade to be blocked but moreover this is the only thing that can stop a cascade.
- To do so, we need a more precise definition.

Definition

A non-empty subset S of nodes is a blocking cluster of density p if every node $v \in S$ has at least a fraction p of its edges go to nodes in S.

Aside

- Clustering is a pervasive concept in many fields and contexts (beyond networks).
- It is an intuitive concept that can be defined in many ways.
- There does not appear to be any one definition that is always (or even usually) most preferred.

Clusters at different levels of granularity

- The given definition of a blocking cluster does not imply a unique way of clustering the nodes.
- Indeed if S and T are both clusters of density p, then the union of S and T is a cluster of density p.
  - Note: this is not generally true of the intersection of S and T.
- This clustering definition also implies that the set of all nodes is a cluster of density 1.

Clusters vs complete cascades

- Suppose we have a network threshold spread model with threshold q, an initial set of A adopters I and $V' = V - I$ is the set of nodes that are not initial adopters.

- Then we have the following (provable) intuitive result that characterizes when complete cascades will or will not occur:

  - If V' contains a cluster C of density greater than $1 - q$, then the initial adopters will not cause a complete cascade. Furthermore, no node in C will adopt A.
  - If in a network with threshold q and an initial set I of adopters does not cause a complete cascade, then the non initial adopters nodes $V' = V - I$ must contain a cluster of density greater than $1 - q$.

When nodes have different thresholds

- As remarked before the assumption that all nodes have the same threshold is not essential.
- Consider a node v. Suppose now that for every adjacent edge (v, w), node v has payoff $a(v)$ (resp. $b(v)$) if both v and w have adopted product A (resp. B) and a zero payoff if v and w currently utilize different products.
- If node v has not yet adopted A at time t, but a fraction p of the $d(v)$ neighbours of v have already adopted A, then:
  - By switching, v has payoff $p \times d(v) \times a(v)$.
  - By staying with B, v has payoff $(1 - p) \times d(v) \times b(v)$.
- Thus node v will switch to A if

$$p \times d(v) \times a(v) \geq (1 - p) \times d(v) \times b(v).$$

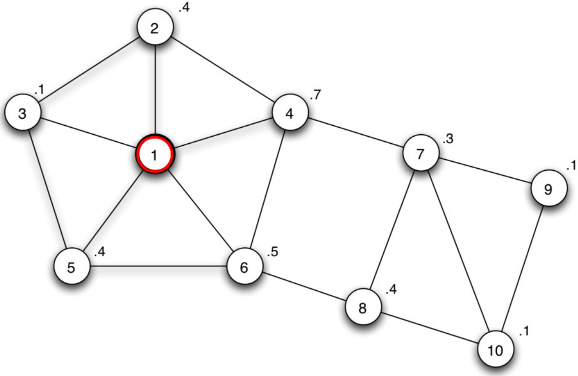
- This is then equivalent to saying that v will switch whenever

$$p \geq \frac{b(v)}{a(v) + b(v)} = q(v)$$

which is then the threshold for node v.

Redefining blocking clusters

- A blocking cluster is now a set of nodes C such that every node $v \in C$ has more than a fraction $1 - q(v)$ of its adjacent nodes in C.
- It follows (as in the case of homogeneous threshold nodes) that a given set of adopters I in a network will not cause a complete cascade iff $V - I$ contains a blocking cluster C.

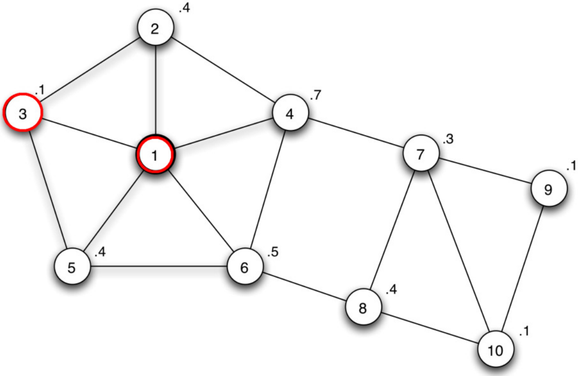


$t = 0$

[Fig 19.13, E&K]

Redefining blocking clusters

- A blocking cluster is now a set of nodes C such that every node $v \in C$ has more than a fraction $1 - q(v)$ of its adjacent nodes in C.
- It follows (as in the case of homogeneous threshold nodes) that a given set of adopters I in a network will not cause a complete cascade iff $V - I$ contains a blocking cluster C.

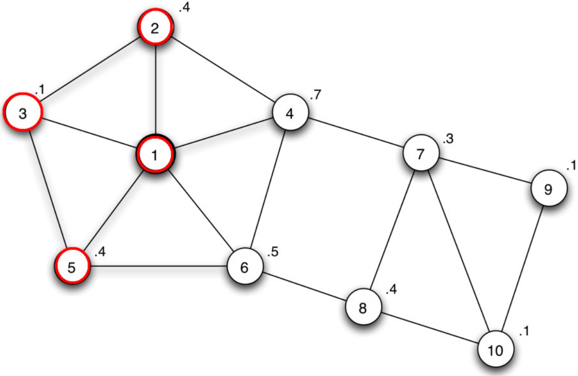


[Fig 19.13, E&K]

$t = 1$

Redefining blocking clusters

- A blocking cluster is now a set of nodes C such that every node $v \in C$ has more than a fraction $1 - q(v)$ of its adjacent nodes in C.
- It follows (as in the case of homogeneous threshold nodes) that a given set of adopters I in a network will not cause a complete cascade iff $V - I$ contains a blocking cluster C.

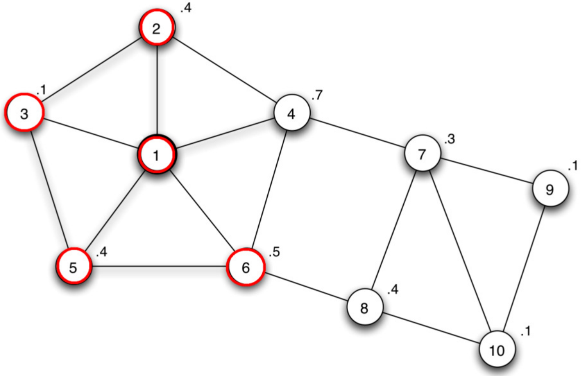


[Fig 19.13, E&K]

$t = 2$

Redefining blocking clusters

- A blocking cluster is now a set of nodes C such that every node $v \in C$ has more than a fraction $1 - q(v)$ of its adjacent nodes in C.
- It follows (as in the case of homogeneous threshold nodes) that a given set of adopters I in a network will not cause a complete cascade iff $V - I$ contains a blocking cluster C.

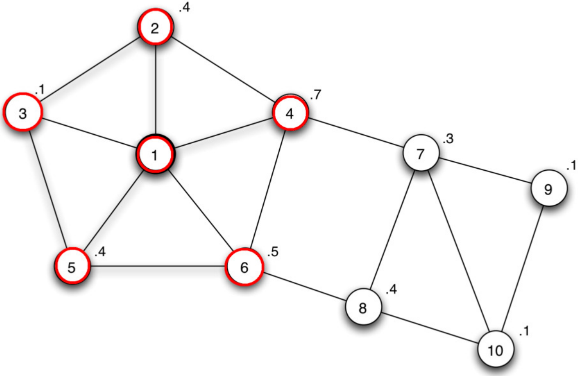


[Fig 19.13, E&K]

$t = 3$

Redefining blocking clusters

- A blocking cluster is now a set of nodes C such that every node $v \in C$ has more than a fraction $1 - q(v)$ of its adjacent nodes in C.
- It follows (as in the case of homogeneous threshold nodes) that a given set of adopters I in a network will not cause a complete cascade iff $V - I$ contains a blocking cluster C.



[Fig 19.13, E&K]

$t = 4$

Further considerations: the "bilingual option"

- In the advanced material (Section 19.7C), the possibility of a third option is considered.
- Here the model allows an individual to maintain both technologies (languages, ideologies, cultural practices) but at a cost c.
- Every individual now can choose to be unilingual (adopting just A or just B) or to be bilingual adopting both (denoted AB).
- The coordination benefit (for each edge) is represented in Figure 19.18. The cost is subtracted from the total benefit over all edges

		w: A	w: B	w: AB
v	A	a, a	$0, 0$	a, a
v	B	$0, 0$	b, b	b, b
v	AB	a, a	b, b	$(a,b)^+, (a,b)^+$

Figure: A Coordination Game with a bilingual option. Here the notation $(a, b)^+$ denotes the larger of a and b. [Fig 19.18, E&K]

Choosing influential adopters

- Suppose we wish to spread a new technology and to do so we have money to influence some "small" set of initial adopters (e.g. by giving away the product or even paying people to adopt it).

- Even in this simple model of (non-competitive) influence spread, and even if we have complete knowledge of the social network, it is not at all clear how to chose an initial set of adopters so as to achieve the largest spread.

- Furthermore the spread process could be much more sophisticated.
  - For example, adoption by a node might be a more random process (say adopting with some probability relative to the nodes threshold) and maybe the influence of neighbors first increases and then decreases over time.

Choosing influential adopters continued

- Suppose we have funds/ability to influence k nodes to become initial adopters.
  - We can try all possible subsets of the entire $n = |V|$ nodes and for each such subset simulate the spread process.
  - But clearly as k gets larger, this "brute force" becomes prohibitive for large (and not even massive) networks.

- It turns out that the problem of the optimum set of initial adopters in many settings is an NP-hard problem.

Can we determine a "good" set of initial adopters?

- For even simple models of information spread similar to those being discussed here, it can be computationally difficult (NP-Hard) to obtain an approximation within a factor n^c for any $c < 1$.

- Instead we will identify properties of a spread process that will allow a good approximation: a good set of initial adopters that will do "almost as well" as the best set.

Note: What follows is a discussion as to how to choose a set of initial adopters by a relatively efficient approximation algorithm when making some assumptions on the spread process. However, we would need much more efficient methods for massive networks.

Influence maximization models; monotone submodular set functions

- Some spread models have the following nice properties.

For any initial set of adopters, S, let $f(S)$ be size (or more generally a real value benefit since some nodes may be more valuable) of the final set of adopters. Furthermore, let f satisfy:

1. Monotonicity: $f(S) \leq f(T)$ if S is a subset of T
2. Submodularity: $f(S+v) - f(S) \geq f(T+v) - f(T)$ if S is a subset of T

- We also usually assume that $f(\emptyset) = 0$. Such normalized, monotone, submodular functions arise in many applications.
- The simple threshold examples considered thus far are monotone processes but are not submodular in general. Are these contrived worst case network examples?
- Some variants of the threshold model and related models do satisfy these properties. Next week, we consider two such stochastic models.

Recap

- Twitter rumour cascades
- Structural virality
- Threshold model
  - Complete cascades
  - Blocking clusters
  - A first look at selecting initial adopters