Analysing information flow in evolving networks using dynamic communicability scores

Billiejoe Charlton
University of Brighton

Acknowledgements

- "In the mood: the dynamics of collective sentiments on Twitter"
  - In Royal Society Open Science Volume 3, Issue 6 (June 2016)
  - Joint work with Colin Singleton and Danica Vukadinovic Greetham
  - Supported by MoD through Centre for Deference Enterprise grant CDE36620
- "Separating temporal and topological effects in walk-based network centrality"
  - In Physics Review E, Volume 94, Issue 1 (July 2016)
  - Joint work with Ewan Colman
- Research done at Counting Lab and University of Reading

Centre for the Mathematics of Human Behaviour

Overview (1/2)

1. background: dynamic communicability in evolving networks
   - a way of measuring how information propagates in an evolving network
   - generalises Katz centrality for (static) graphs
   - developed by Grindrod, Higham, Estrada and others
   - assigns *broadcast scores* to nodes, quantifying their “communication reach”

2. dynamic communicability and sentiment in social networks
   - we studied dynamic communicability and sentiment in a large sample of Twitter data
   - we found that people with the highest communication reach use more positive sentiment
   - and less negative sentiment than the average user

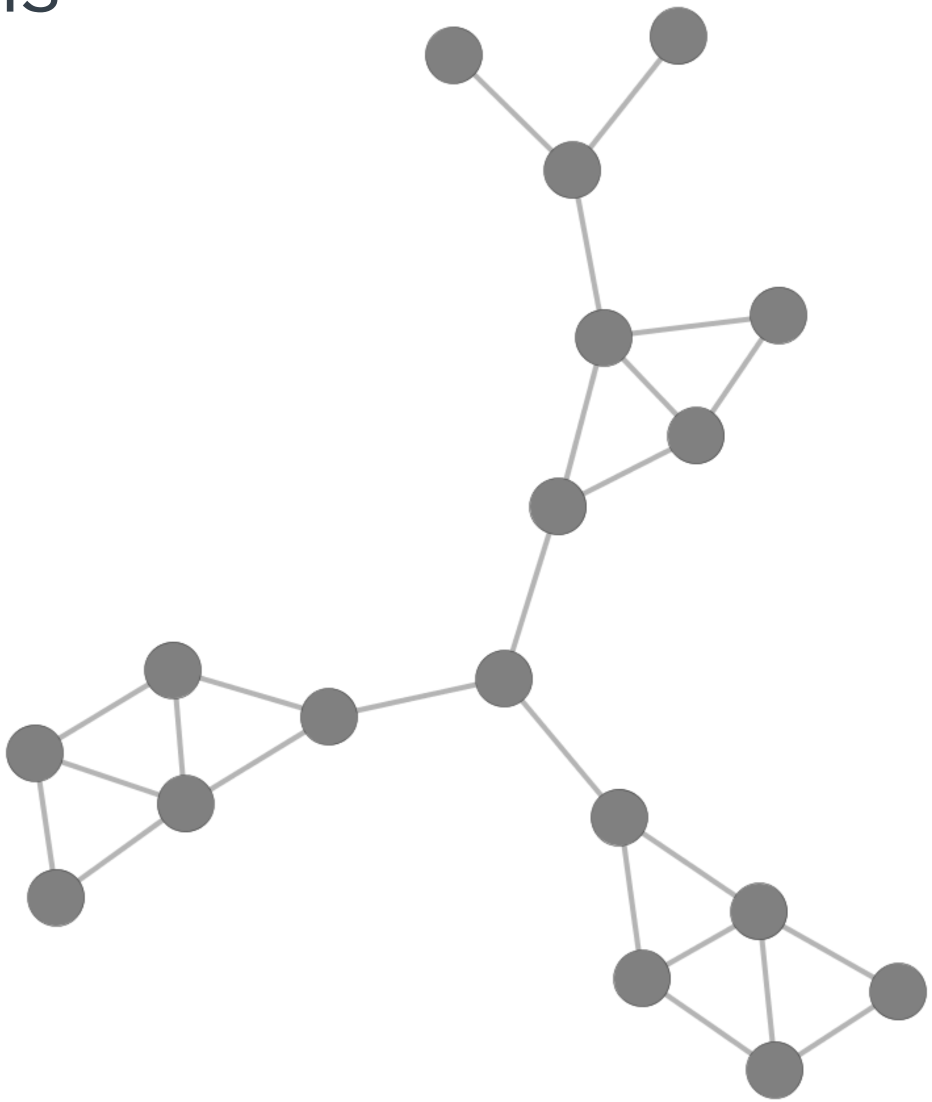
Overview (2/2)

3. separating volume, topological and temporal factors
   - three things affect a node's broadcast score: the volume (number) of messages the node sends, who the messages are sent to, (topological) and when they are sent (temporal)
   - but the broadcast score just gives you a single number
   - this is a bit frustrating
   - we showed a way to separate the effects of the three factors
   - solution requires solving appropriate matrix different equations

Dynamic communicability in evolving networks

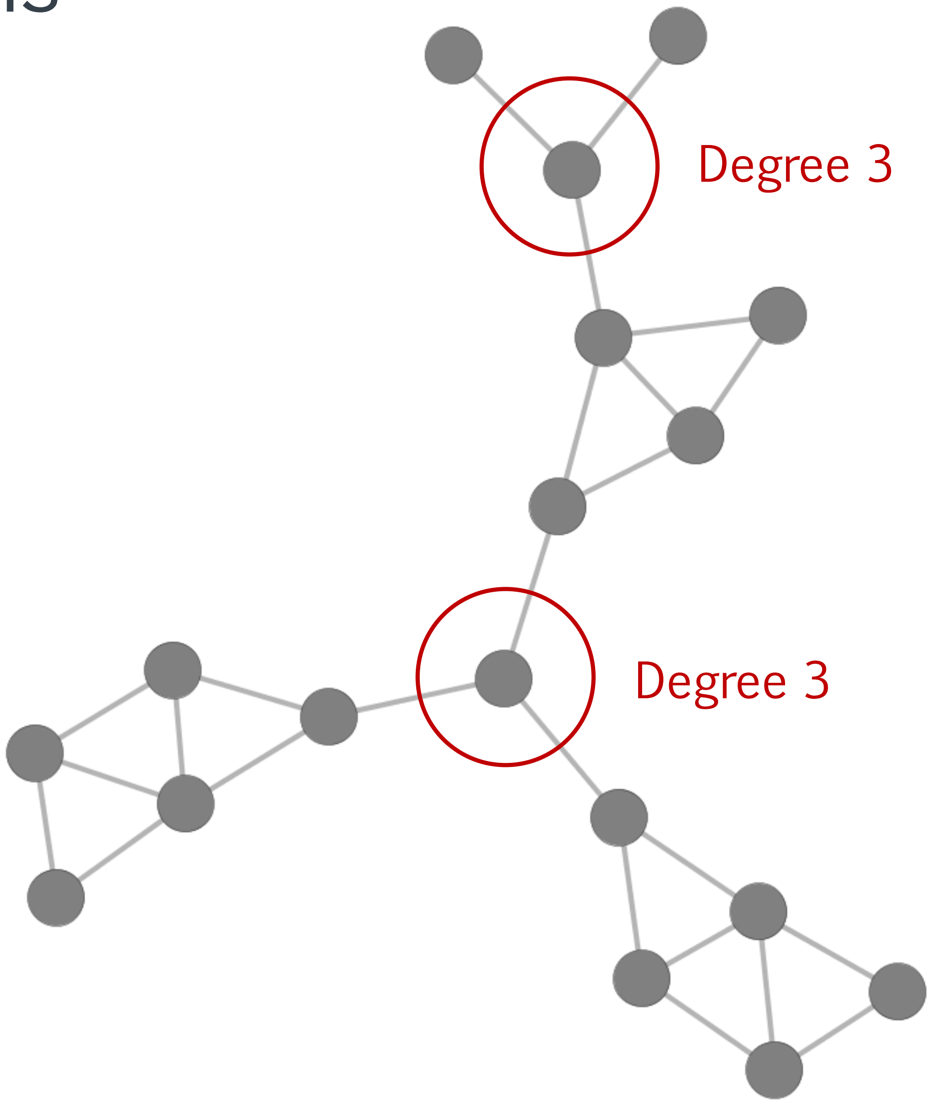
Katz Centrality in static graphs

- Suppose we have a graph representing a social network
- Which people (nodes) are most important? most influential?



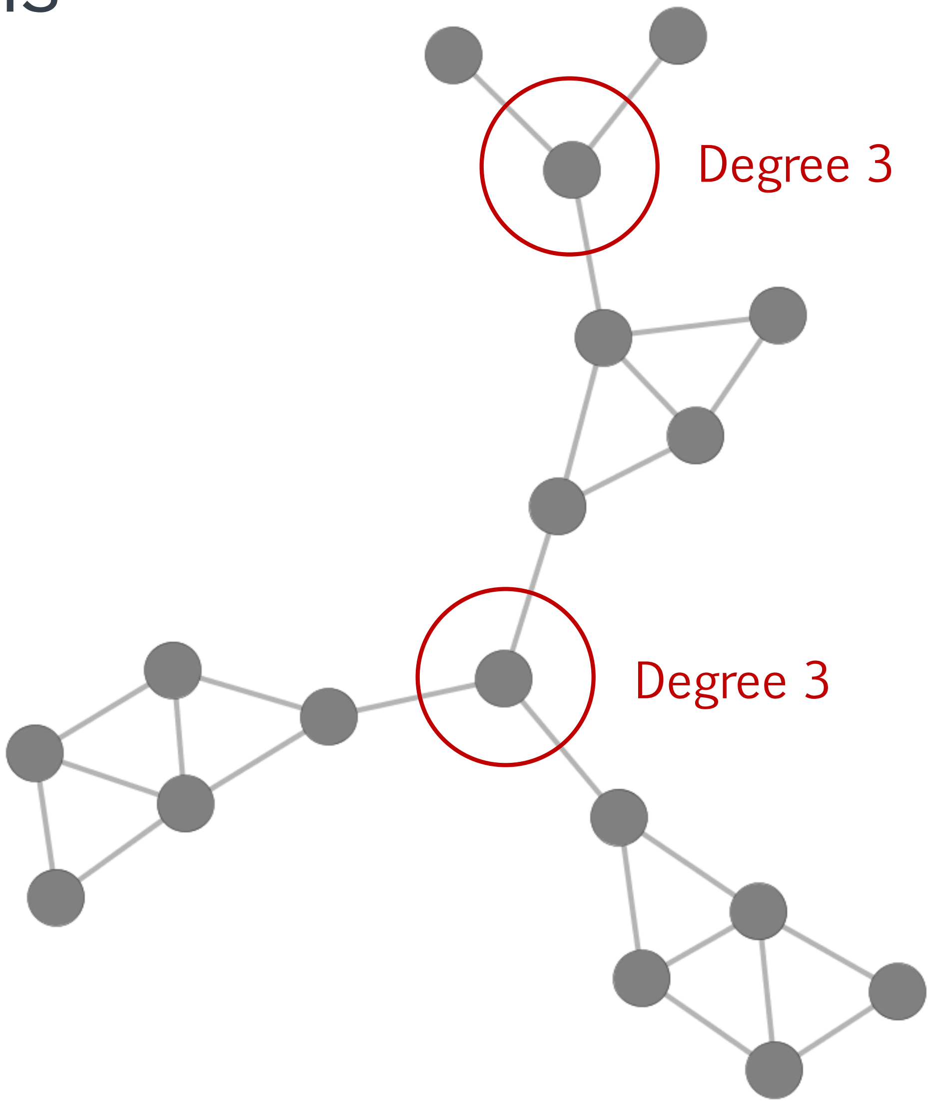
Katz Centrality in static graphs

- Suppose we have a graph representing a social network
- Which people (nodes) are most important? most influential?
- Degree doesn't work very well
- Network structure matters!
  - timing will matter too



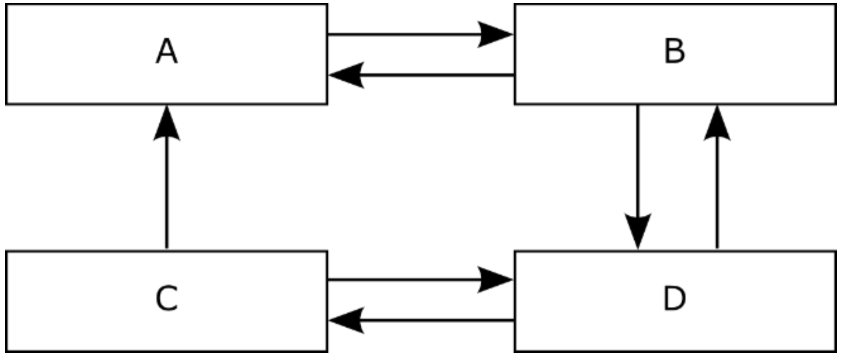
Katz Centrality in static graphs

- Suppose we have a graph representing a social network
- Which people (nodes) are most important? most influential?
- Degree doesn't work very well
- Network structure matters!
  - timing will matter too
- **Katz Centrality** is an early attempt to turn network structure into a number reflecting the centrality / "influence" / "importance" of each node
  - L. Katz, "A New Status Index Derived from Sociometric Index" (1953).



Katz Centrality in static graphs

› For illustration let's take a simpler graph:

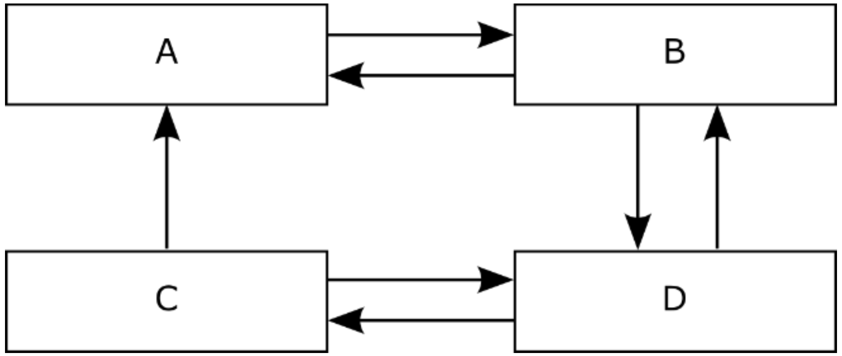


› Represent it as an adjacency matrix:

$$\begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:

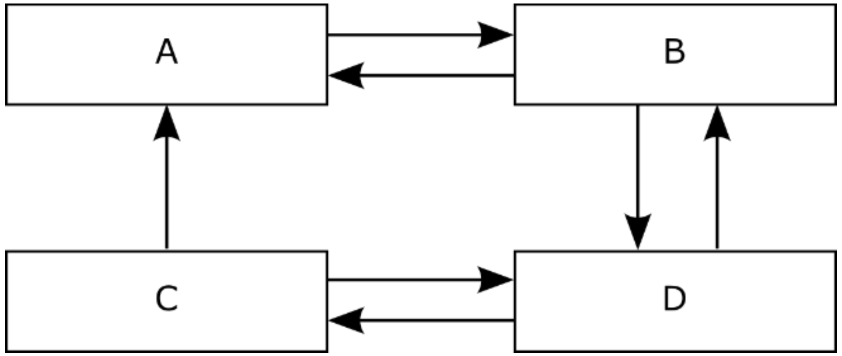


- **The squared matrix counts paths of length two**

$$A^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} \end{array}$$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:

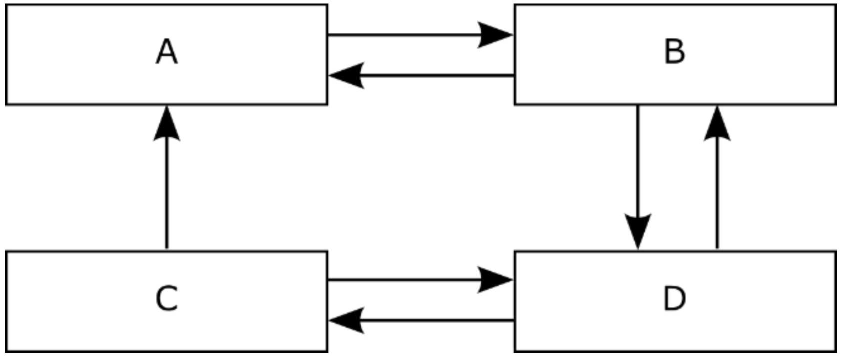


- **The squared matrix counts paths of length two**

$$A^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \text{Source} \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} & \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} \end{array}$$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:

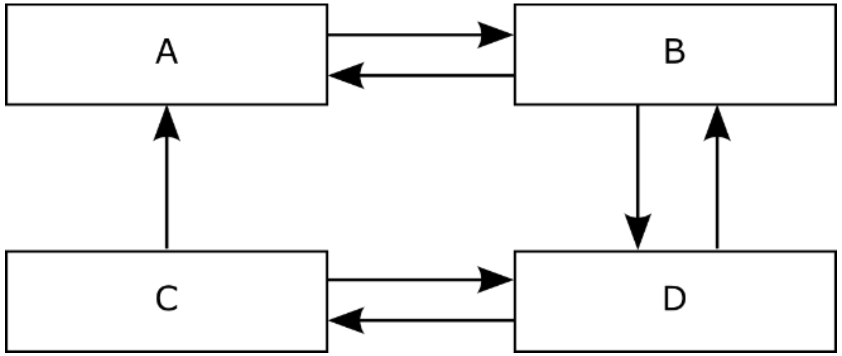


- **The cubed matrix counts paths of length three**

$$A^3 = \begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 0 & 2 & 1 & 0 \\ 3 & 0 & 0 & 3 \\ 3 & 0 & 0 & 3 \\ 0 & 4 & 2 & 0 \end{pmatrix} \end{array}$$

Katz Centrality in static graphs

› For illustration let's take a simpler graph:



› **The cubed matrix counts paths of length three**

Paths:
- D → C → A → B
- D → C → D → B
- D → B → A → B
- D → B → D → B

$$A^3 = \begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 0 & 2 & 1 & 0 \\ 3 & 0 & 0 & 3 \\ 3 & 0 & 0 & 3 \\ 0 & 4 & 2 & 0 \end{pmatrix} \end{array}$$

Katz Centrality formula

- In general, the entries of A^n count the number of paths of length n
- Katz' idea is to look at the matrix

$$C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k$$

Katz Centrality formula

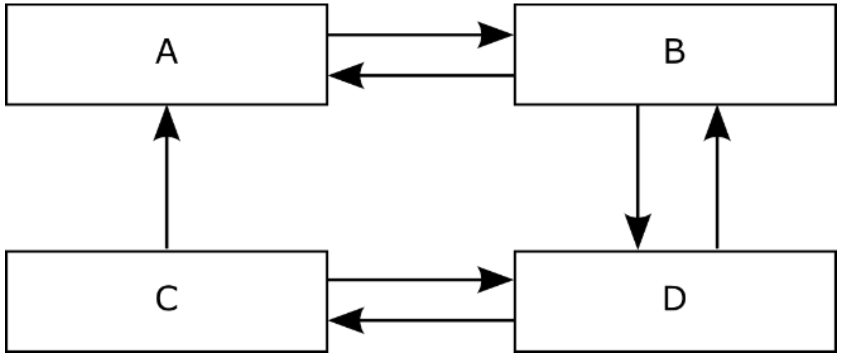
- In general, the entries of A^n count the number of paths of length n
- Katz' idea is to look at the matrix

$$C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k$$

- i.e. **combine the information about reachability by all possible lengths of path**
- Here α is for down-weighting longer paths
  - direction connections are more influential than indirect paths
- For convergence we need $\alpha < 1/\rho(A)$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:

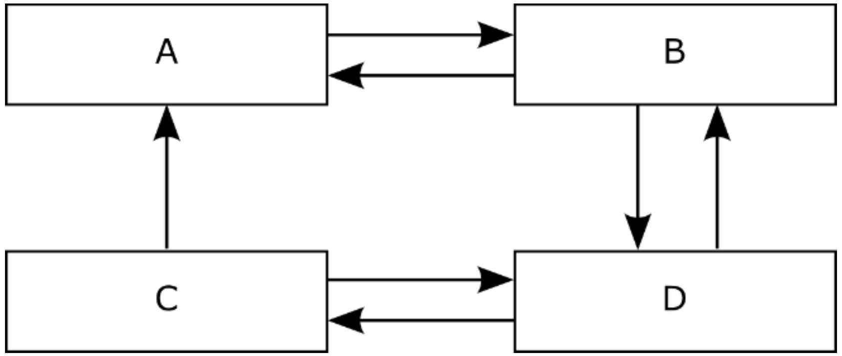


- For the above graph with $\alpha = 0.5$ we get:

$$C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k = I + \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} + \frac{1}{8}\begin{pmatrix} 0 & 2 & 1 & 0 \\ 3 & 0 & 0 & 3 \\ 3 & 0 & 0 & 3 \\ 0 & 4 & 2 & 0 \end{pmatrix} + \dots$$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:



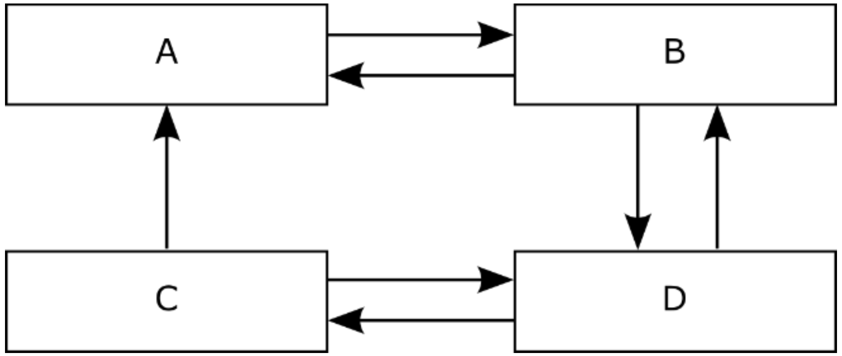
- For the above graph with $\alpha = 0.5$ we get:

$$C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k = I + \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} + \frac{1}{8}\begin{pmatrix} 0 & 2 & 1 & 0 \\ 3 & 0 & 0 & 3 \\ 3 & 0 & 0 & 3 \\ 0 & 4 & 2 & 0 \end{pmatrix} + \ldots$$

E.g. getting from node D to node B: $\frac{1}{2}.1 + \frac{1}{4}.0 + \frac{1}{8}.4 + \ldots$

Katz Centrality in static graphs

- For illustration let's take a simpler graph:

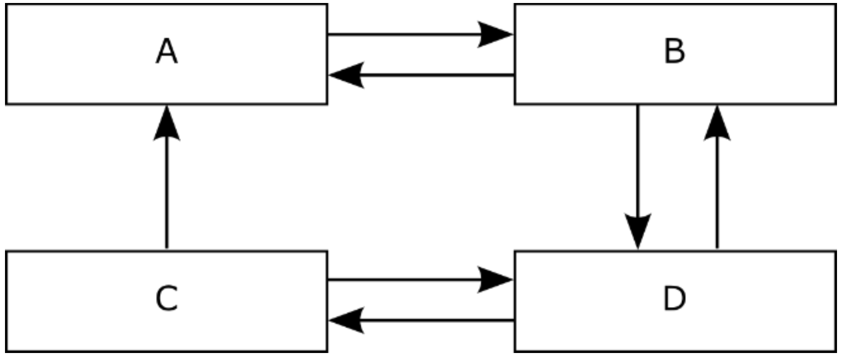


- For the above graph with $\alpha = 0.5$ we get:

$$C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k = \begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 4.27 & 4.15 & 1.80 & 3.27 \\ 5.95 & 7.54 & 3.27 & 5.95 \\ 5.95 & 6.54 & 4.27 & 5.95 \\ 6.54 & 7.74 & 4.15 & 7.74 \end{pmatrix} \end{array}$$

Katz Centrality in static graphs

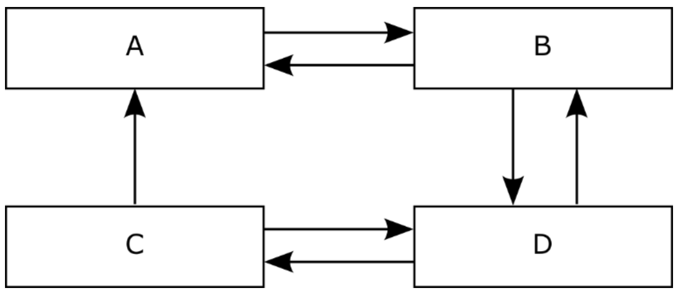
- For illustration let’s take a simpler graph:



- For the above graph with $\alpha = 0.5$ we get:

$$
C_{Katz} = \sum_{k=0}^{\infty} \alpha^k A^k = \begin{array}{c c} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 4.27 & 4.15 & 1.80 & 3.27 \\ 5.95 & 7.54 & 3.27 & 5.95 \\ 5.95 & 6.54 & 4.27 & 5.95 \\ 6.54 & 7.74 & 4.15 & 7.74 \end{pmatrix} \end{array}
$$

Broadcast and receive scores



› Row and column sums have been used as “broadcast scores” and “receive scores”

Target

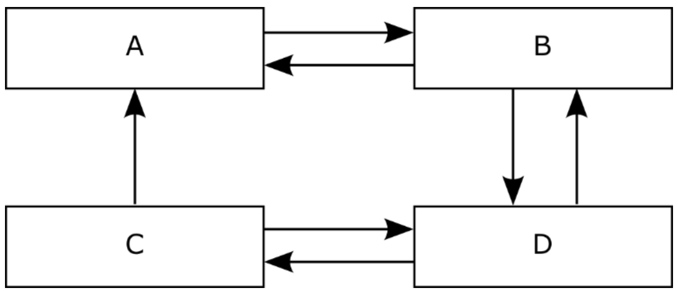
$$
C_{Katz} = \begin{array}{cc} & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source}\begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 4.27 & 4.15 & 1.80 & 3.27 \\ 5.95 & 7.54 & 3.27 & 5.95 \\ 5.95 & 6.54 & 4.27 & 5.95 \\ 6.54 & 7.74 & 4.15 & 7.74 \end{pmatrix} \end{array} \qquad \begin{pmatrix} 13.5 \\ 22.7 \\ 22.7 \\ 25.9 \end{pmatrix}
$$

Row sums – “broadcast scores”

$$
\begin{pmatrix} 22.7 & 25.9 & 13.5 & 22.7 \end{pmatrix}
$$

Column sums – “receive scores”

Broadcast and receive scores



› Row and column sums have been used as "broadcast scores" and "receive scores"

$$
\begin{array}{cc}
 & \text{Target} \\
 & \begin{array}{cccc} A & B & C & D \end{array} \\
C_{Katz} = \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{pmatrix} 4.27 & 4.15 & 1.80 & 3.27 \\ 5.95 & 7.54 & 3.27 & 5.95 \\ 5.95 & 6.54 & 4.27 & 5.95 \\ 6.54 & 7.74 & 4.15 & 7.74 \end{pmatrix}
\end{array}
\qquad
\begin{pmatrix} 13.5 \\ 22.7 \\ 22.7 \\ 25.9 \end{pmatrix}
$$

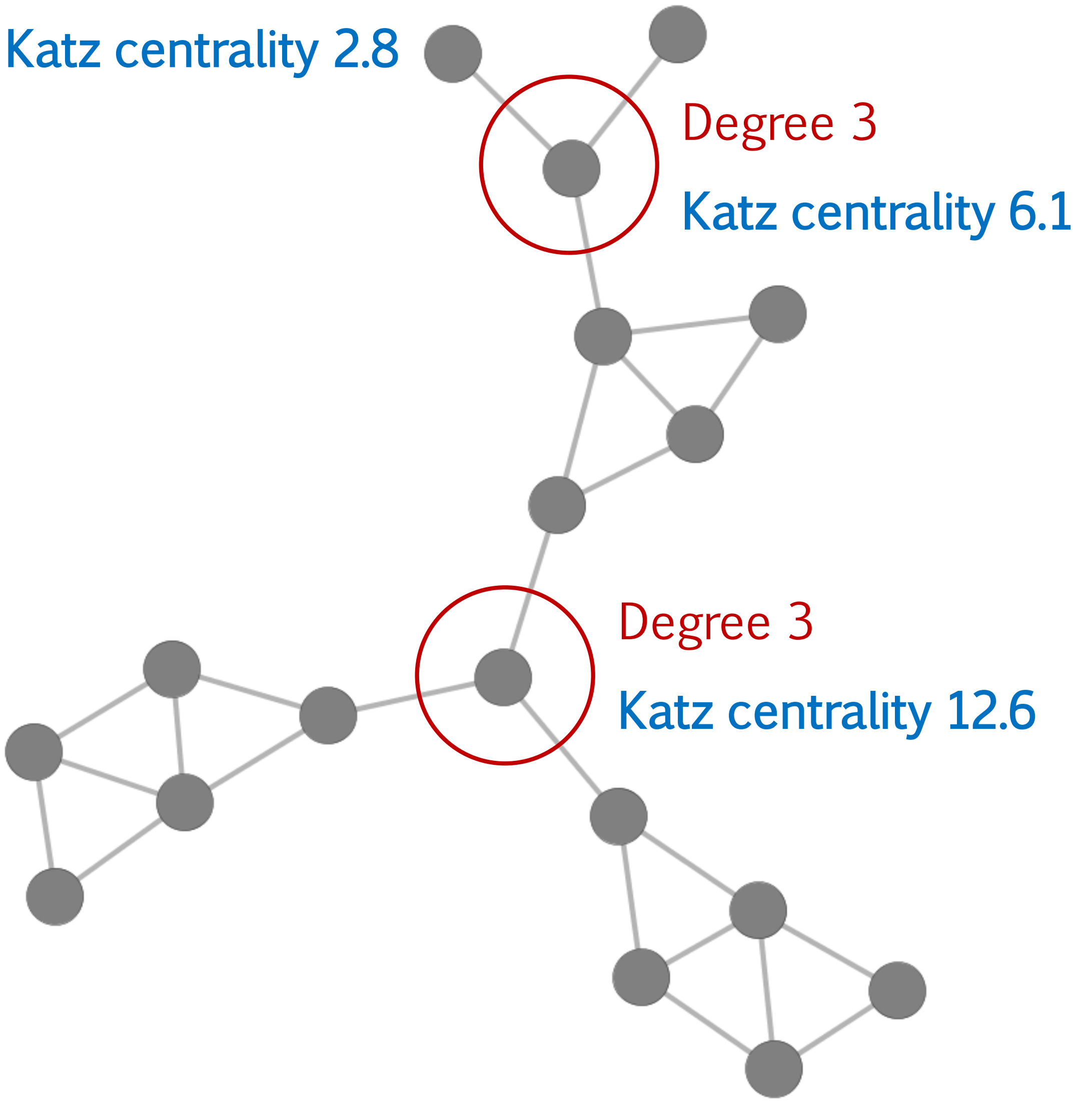
Row sums – "broadcast scores"

$$
\begin{pmatrix} 22.7 & 25.9 & 13.5 & 22.7 \end{pmatrix}
$$

Column sums – "receive scores"

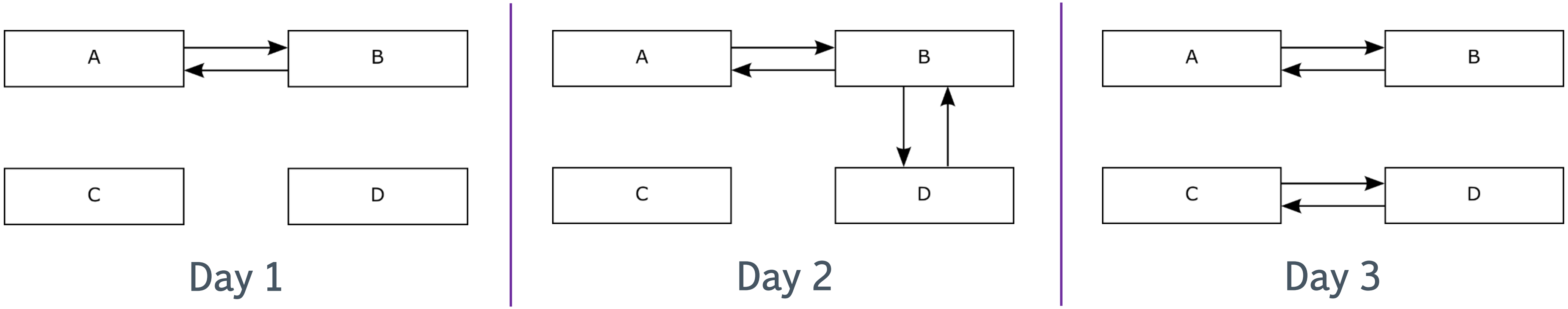
Katz Centrality in static graphs

With $\alpha = 0.3$



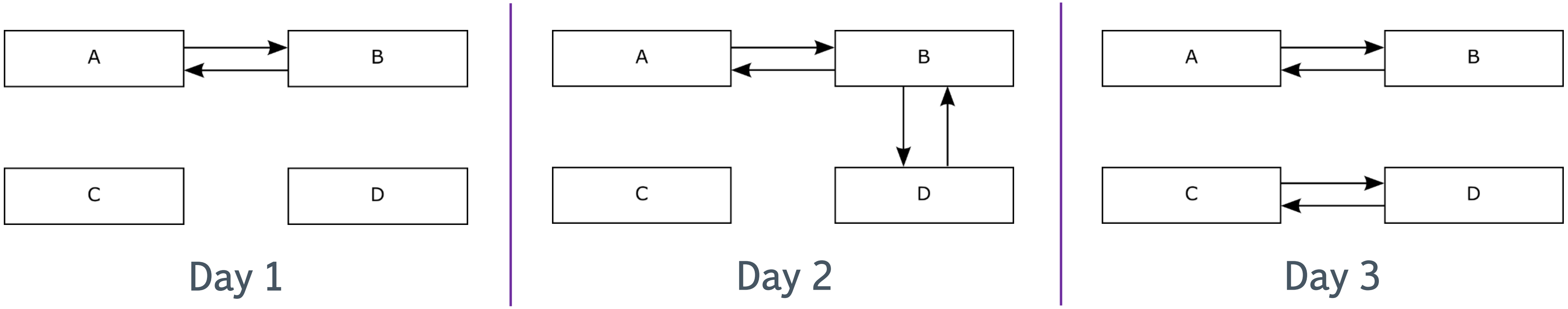
Now let's add time...

- Instead of one network, we have a sequence of successive snapshots $A_{(1)}, A_{(2)}, \ldots, A_{(M)}$



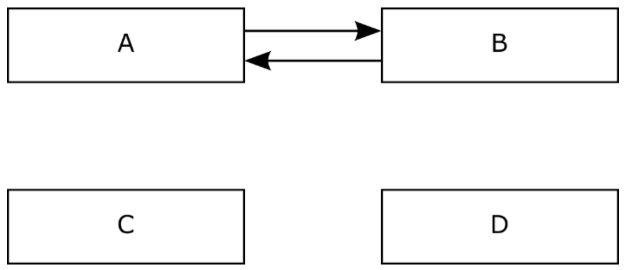
Now let's add time…

- Instead of one network, we have a sequence of successive snapshots $A_{(1)}, A_{(2)}, \dots, A_{(M)}$

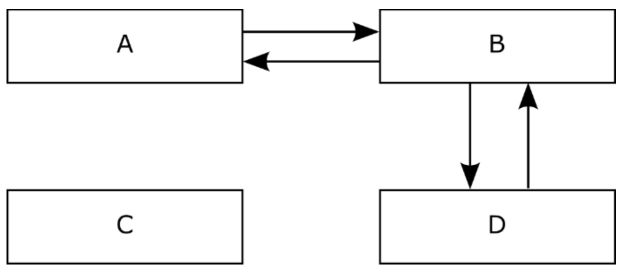


We then define the **dynamic communicability matrix** as

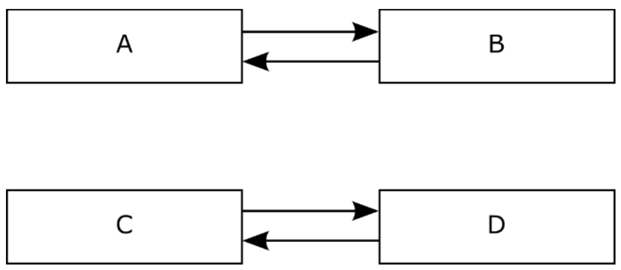
$$Q = \prod_{i=1}^{M} \underbrace{\sum_{k=0}^{\infty} \alpha^k {A_{(i)}}^k}_{\text{This is just the Katz centrality matrix for time } i}$$



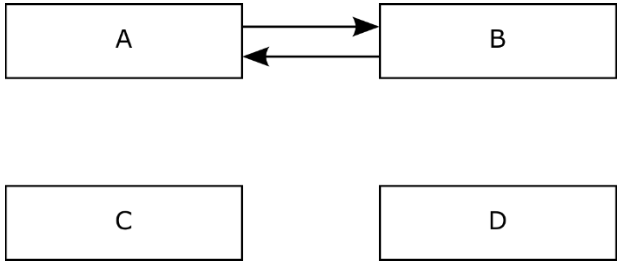
Day 1



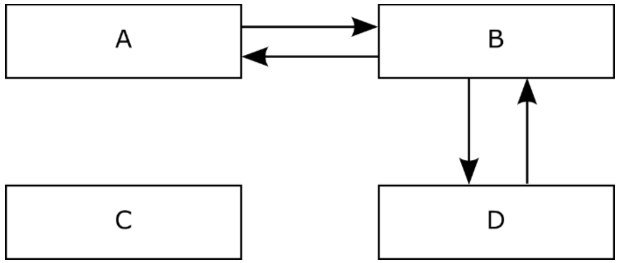
Day 2



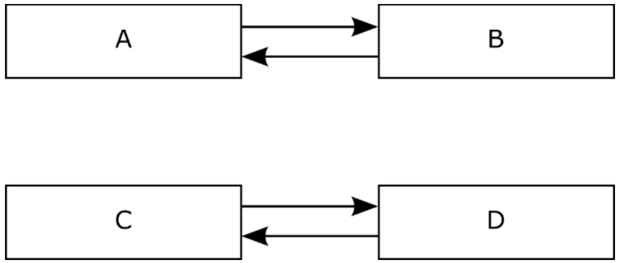
Day 3



Day 1



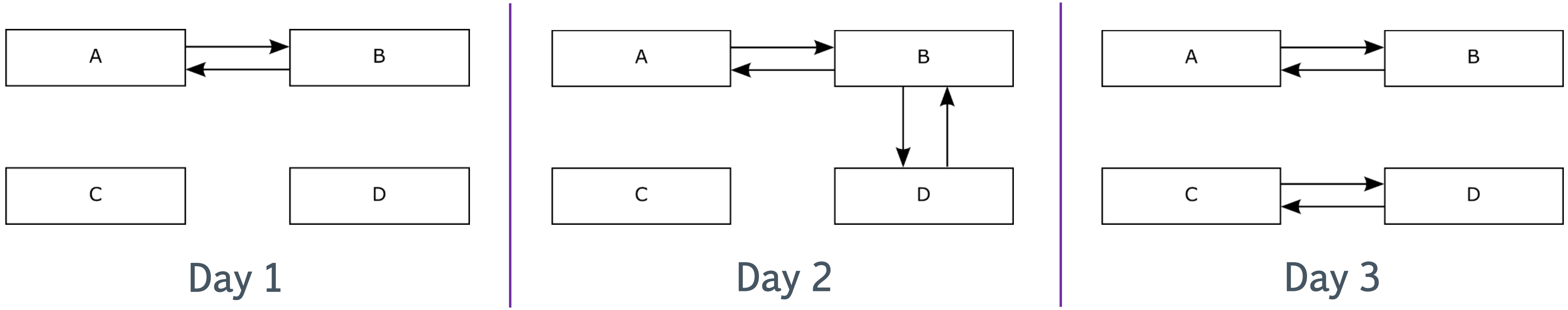
Day 2



Day 3

- Respective Katz centrality matrices (with $\alpha = 0.3$):

$$\frac{1}{3}\begin{pmatrix}4 & 2 & 0 & 0\\ 2 & 4 & 0 & 0\\ 0 & 0 & 3 & 0\\ 0 & 0 & 0 & 3\end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix}3 & 2 & 0 & 1\\ 2 & 4 & 0 & 2\\ 0 & 0 & 2 & 0\\ 1 & 2 & 0 & 3\end{pmatrix} \qquad \frac{1}{3}\begin{pmatrix}4 & 2 & 0 & 0\\ 2 & 4 & 0 & 0\\ 0 & 0 & 4 & 2\\ 0 & 0 & 2 & 4\end{pmatrix}$$



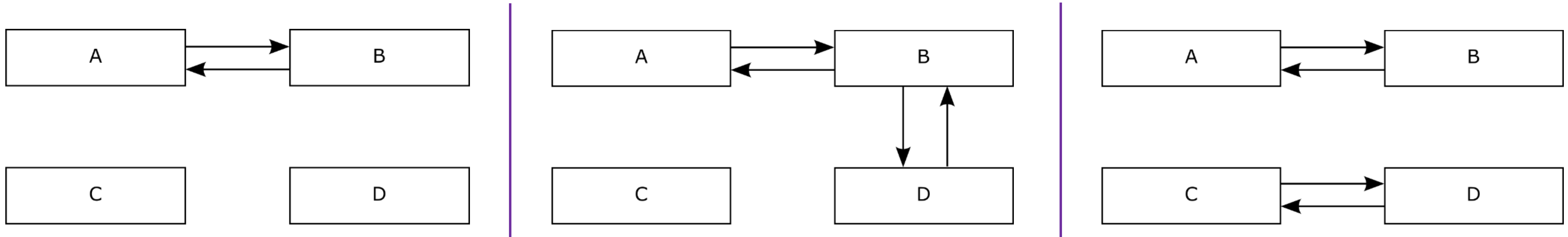
- Respective Katz centrality matrices (with $\alpha = 0.3$):

$$\frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 3 & 2 & 0 & 1 \\ 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix} \qquad \frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

- Communicability matrix Q:

- Effectively we are counting **“time-respecting walks”**, and down-weighting for length

$$\begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \dfrac{1}{9}\begin{pmatrix} 48 & 48 & 8 & 16 \\ 48 & 54 & 10 & 20 \\ 0 & 0 & 12 & 6 \\ 12 & 15 & 9 & 18 \end{pmatrix} \end{array}$$



Day 1 | Day 2 | Day 3

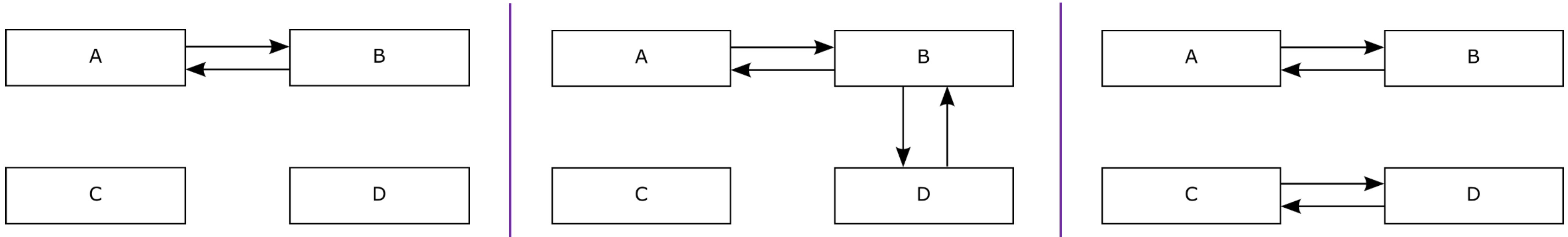
- Respective Katz centrality matrices (with $\alpha = 0.3$):

$$\frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 3 & 2 & 0 & 1 \\ 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix} \qquad \frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

- Communicability matrix Q:

- Effectively we are counting **"time-respecting walks"**, and down-weighting for length

$$\begin{array}{cc} & \text{Target} \\ & \begin{array}{cccc} A & B & C & D \end{array} \\ \text{Source} \begin{array}{c} A \\ B \\ C \\ D \end{array} & \frac{1}{9}\begin{pmatrix} 48 & 48 & 8 & 16 \\ 48 & 54 & 10 & 20 \\ 0 & 0 & 12 & 6 \\ 12 & 15 & 9 & 18 \end{pmatrix} \end{array}$$



Day 1 | Day 2 | Day 3

› Respective Katz centrality matrices (with $\alpha = 0.3$):

$$\frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 3 & 2 & 0 & 1 \\ 2 & 4 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix} \qquad \frac{1}{3}\begin{pmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

› Communicability matrix Q:

› Effectively we are counting **“time-respecting walks”**, and down-weighting for length

Target

Row sums - “broadcast scores”

$$\text{Source}\quad \begin{array}{c|c} & \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} & \frac{1}{9}\begin{pmatrix} 48 & 48 & 8 & 16 \\ 48 & 54 & 10 & 20 \\ 0 & 0 & 12 & 6 \\ 12 & 15 & 9 & 18 \end{pmatrix} \end{array} \qquad \frac{1}{54}\begin{pmatrix} 20 \\ 22 \\ 3 \\ 9 \end{pmatrix}$$

Interpretation of broadcast scores

Researchers have made various interpretations of broadcast scores, in different settings:

- In social networks, broadcast scores have been interpreted as quantifying the **level of influence** or **communication reach** of each person

  (this is why we got MoD funding for our work)

- In the context of epidemic spread, where edges represent contact between individuals, broadcast scores have been interpreted as **ability to infect the population**

- In studies of fMRI data ("videos" of the brain) broadcast scores have been interpreted as showing the **importance of different brain regions**

Some comments

- The idea is very simple, but lots of papers are appearing about it nevertheless. E.g.:
  - Discovering and validating influence in a dynamic online social network (Laflin et al). Social Network Analysis and Mining, 3(4), Dec 2013.
  - Dynamic communicability and epidemic spread: a case study on an empirical dynamic contact network (Chen, Benzi, Chang and Hertzberg). J. of Complex Networks, 2016.
  - Dynamic Communicability Predicts Infectiousness (Mantzaris and Higham). In "Temporal Networks“ (Understanding Complex Systems series), Springer, 2013.
  - Dynamic network centrality summarizes learning in the human brain (Mantzaris et al). Journal of Complex Networks, 2013.

Some comments

- The idea is very simple, but lots of papers are appearing about it nevertheless. E.g.:
  - Discovering and validating influence in a dynamic online social network (Laflin et al). Social Network Analysis and Mining, 3(4), Dec 2013.
  - Dynamic communicability and epidemic spread: a case study on an empirical dynamic contact network (Chen, Benzi, Chang and Hertzberg). J. of Complex Networks, 2016.
  - Dynamic Communicability Predicts Infectiousness (Mantzaris and Higham). In "Temporal Networks" (Understanding Complex Systems series), Springer, 2013.
  - Dynamic network centrality summarizes learning in the human brain (Mantzaris et al). Journal of Complex Networks, 2013.

- Dynamic communicability is not magic!
  - but does seem to reveal interesting things about evolving networks

Dynamic communicability and sentiment in social networks

Dynamic communicability and social networks

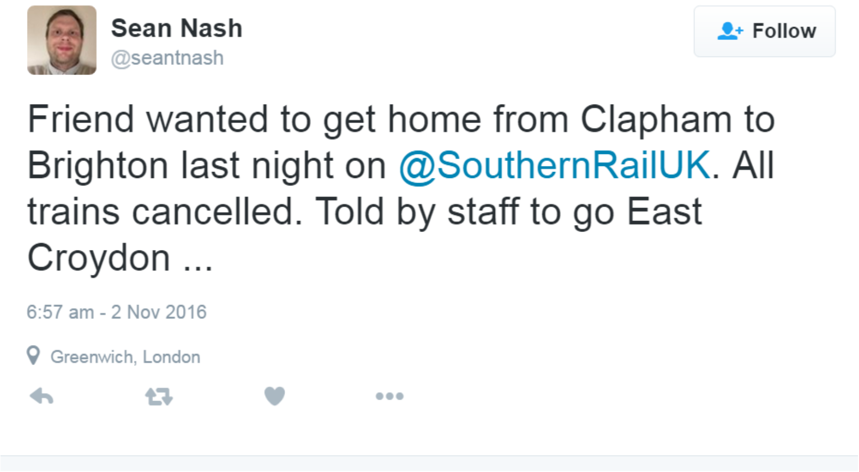
- So we can use broadcast scores to identify influential people in social networks.
- In addition to merely identifying the influential users, we can **investigate how they behave**
  - We can compare them to the average user, and ask **"What is different about them?"**
- In a study* conducted with small face-to-face friendship networks my colleague noticed:

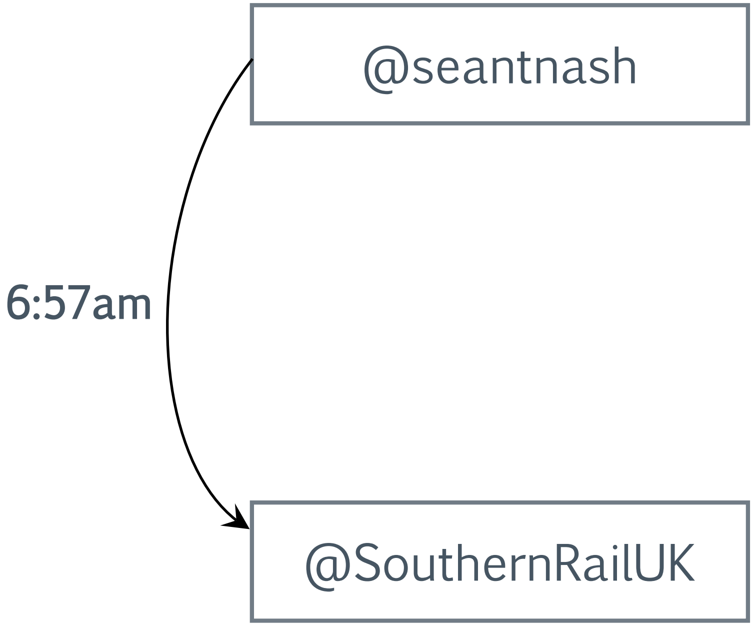
  The people with the highest broadcast scores expressed very little negative emotion
- We set out to see if we could replicate this finding on a large scale
  - using **data from an online social network**
  - and using **sentiment analysis software** to assess positive and negative emotion

* Interventions in social networks: impact on mood and network dynamics (Vukadinovic Greetham, Sengupta, Hurling and Wilkinson). Advances in Complex Systems 18, 2015.

Collecting data

- We collected data from the microblogging platform Twitter
- @-mentions were used to obtain a directed graph with timestamped edges


Sean Nash
@seantnash
Follow
Friend wanted to get home from Clapham to Brighton last night on @SouthernRailUK. All trains cancelled. Told by staff to go East Croydon ...
6:57 am - 2 Nov 2016
Greenwich, London


@seantnash
6:57am
@SouthernRailUK

Sean Nash
@seantnash

Follow

Friend wanted to get home from Clapham to Brighton last night on @SouthernRailUK. All trains cancelled. Told by staff to go East Croydon ...

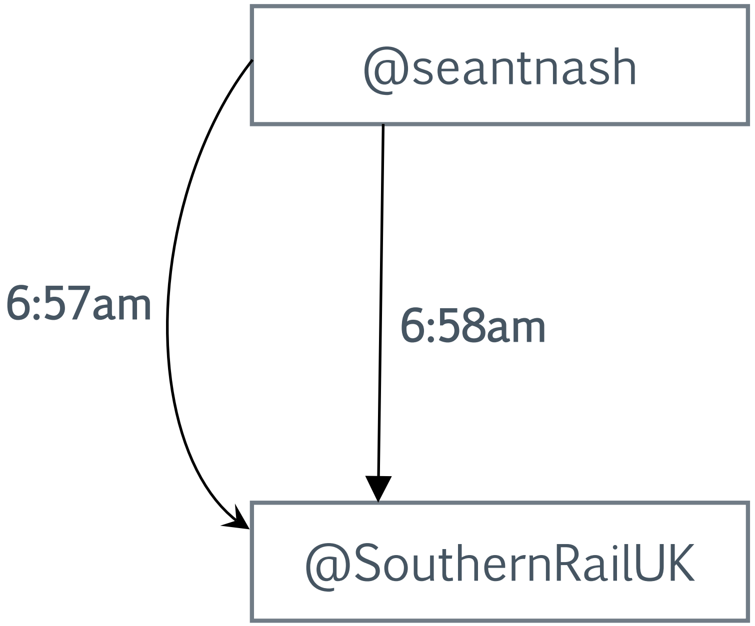
6:57 am - 2 Nov 2016

Greenwich, London

.... and change there. Next train from East Croydon to Brighton was 4:30am - a 5 hour wait!!! Totally disgusting @SouthernRailUK

6:58 am - 2 Nov 2016

Greenwich, London



Sean Nash
@seantnash

Follow

Friend wanted to get home from Clapham to Brighton last night on @SouthernRailUK. All trains cancelled. Told by staff to go East Croydon ...

6:57 am - 2 Nov 2016

Greenwich, London

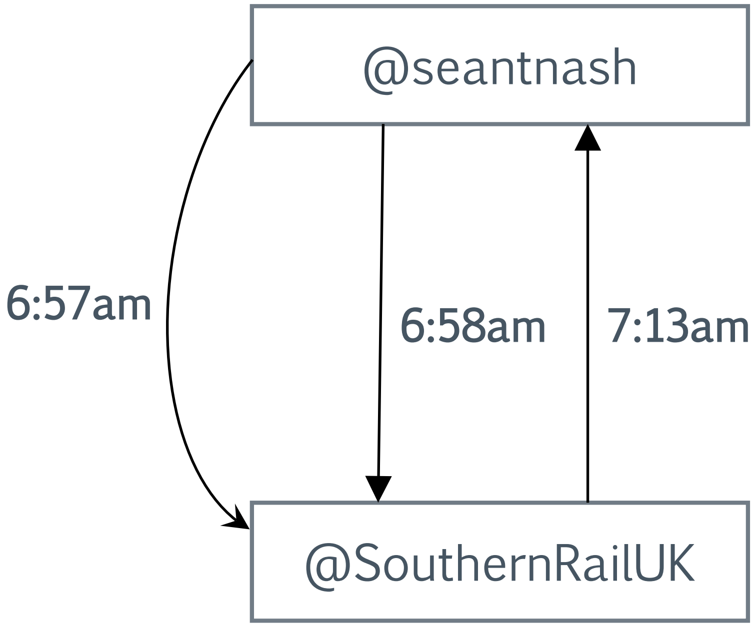
.... and change there. Next train from East Croydon to Brighton was 4:30am - a 5 hour wait!!! Totally disgusting @SouthernRailUK

6:58 am - 2 Nov 2016

Greenwich, London

Southern @SouthernRailUK · Nov 2

@seantnash Did he make it home Sean? T



Sean Nash @seantnash

Follow

Friend wanted to get home from Clapham to Brighton last night on @SouthernRailUK. All trains cancelled. Told by staff to go East Croydon ...

6:57 am - 2 Nov 2016

Greenwich, London

.... and change there. Next train from East Croydon to Brighton was 4:30am - a 5 hour wait!!! Totally disgusting @SouthernRailUK

6:58 am - 2 Nov 2016

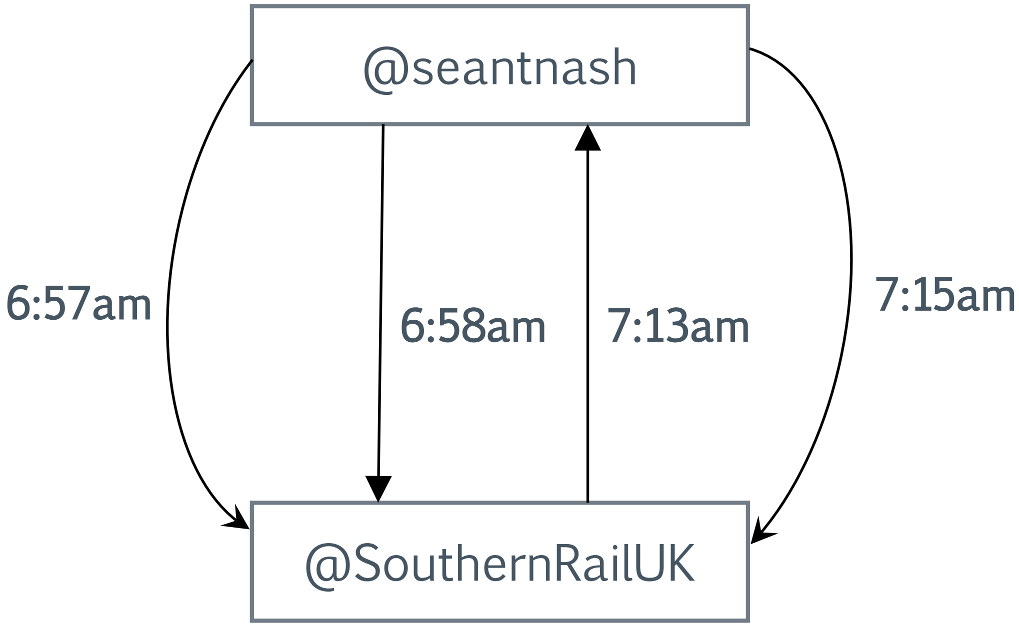
Greenwich, London

Southern @SouthernRailUK · Nov 2

@seantnash Did he make it home Sean? T

Sean Nash @seantnash · Nov 2

@SouthernRailUK yes but only because I offered to drive him home so he wasn't stranded. A 140 mile round trip at midnight. Not good!



Collecting data

- We collected data from the microblogging platform Twitter
- @-mentions were used to obtain a directed graph with timestamped edges

- Snowball sampling was used to select users
- We collected 147 million tweets for ~670k users, mainly covering April 2014 – April 2015

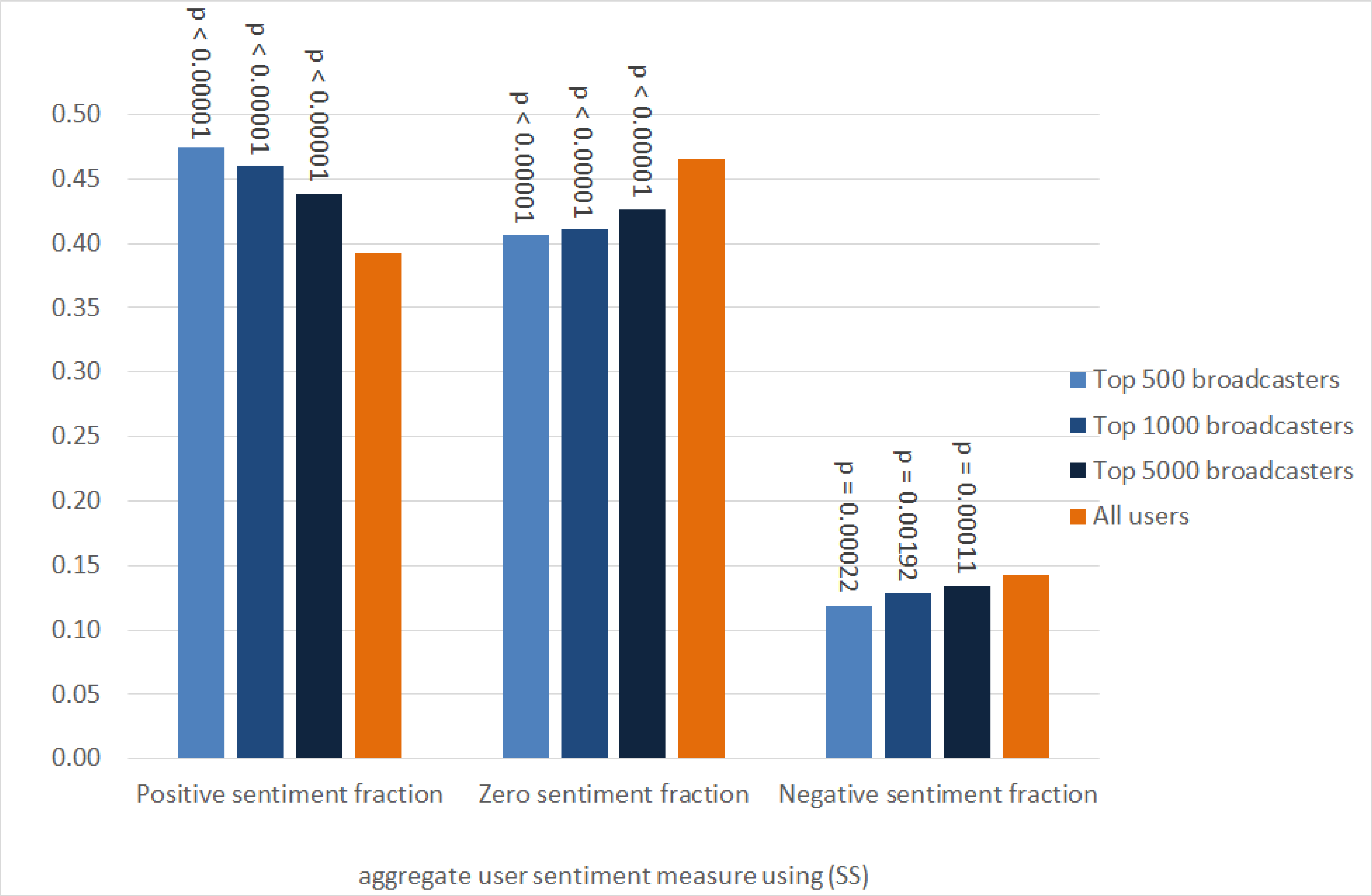
- For this exercise we selected a week period beginning 9th Oct 2014
  - here we had complete data for the largest amount of users
  - after bot removal we had ~6 million edges between ~285k users

- We calculated a **broadcast score** and an **average sentiment** for each user
  - then we look at how these were related

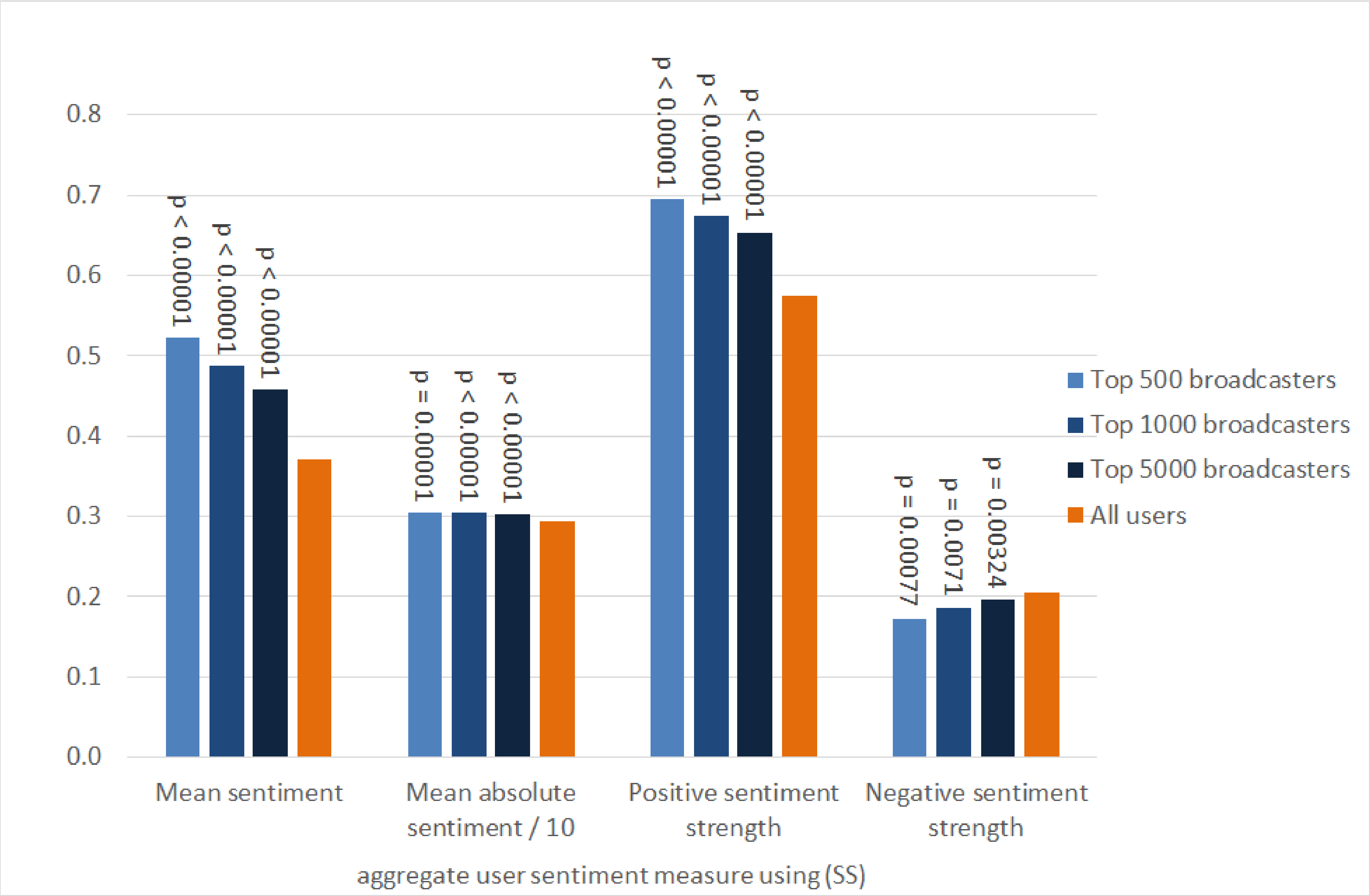
Sentiment scores

Tweet content	Sentiment (-25 to +25)
RT @Otto_English: UKIP attracts the old, the scared, the wrong, the bitter, the racist, the disillusioned, the angry and the stupid. That …	-15
.@biGsAm @RoyalJordanian usual shit. RJ is shit. Move to other less shitty airline, no shit. #fail #amman #jo	-10
@DanHemming not gonna lie, underwhelmed at the moment... But it is only episode one. I'm bearing with.	-2
@Rowetta @katherinemills probably seen photos	0
@MarionManton @gillferrell cool	1
So happy to return to my favourite spot: Cafe Gray Deluxe @UpperHouse_HKG. Yummy dinner with @LinkedInMktg HK team! http://t.co/QK108uN1Qp	5
A huge HUGE Happy Birthday to the brilliant and hilarious @JenSmith1850 - one of my first Twitter friends. Love you pal, have an amazing day	15
They are actual scumbags, and I don't use that word lightly	0
Urgh! Turned on the TV and there was Katie Hopkins' porcine face. Boak!	0

Results



p < 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p = 0.00022
p = 0.00192
p = 0.00011
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Positive sentiment fraction
Zero sentiment fraction
Negative sentiment fraction
Top 500 broadcasters
Top 1000 broadcasters
Top 5000 broadcasters
All users
aggregate user sentiment measure using (SS)



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
p < 0.00001
p < 0.00001
p < 0.00001
p = 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p < 0.00001
p = 0.00077
p = 0.0071
p = 0.00324
Top 500 broadcasters
Top 1000 broadcasters
Top 5000 broadcasters
All users
Mean sentiment
Mean absolute sentiment / 10
Positive sentiment strength
Negative sentiment strength
aggregate user sentiment measure using (SS)

Great! Right?

- We had succeeded in showing, in a large online social network, a relationship between broadcast score and sentiment use.

- It made a journal paper, accompanied by findings about community detection, sentiment and cohesion over time.

Great! Right?

- We had succeeded in showing, in a large online social network, a relationship between broadcast score and sentiment use.
- It made a journal paper, accompanied by findings about community detection, sentiment and cohesion over time.
- But somehow I didn't quite like it.

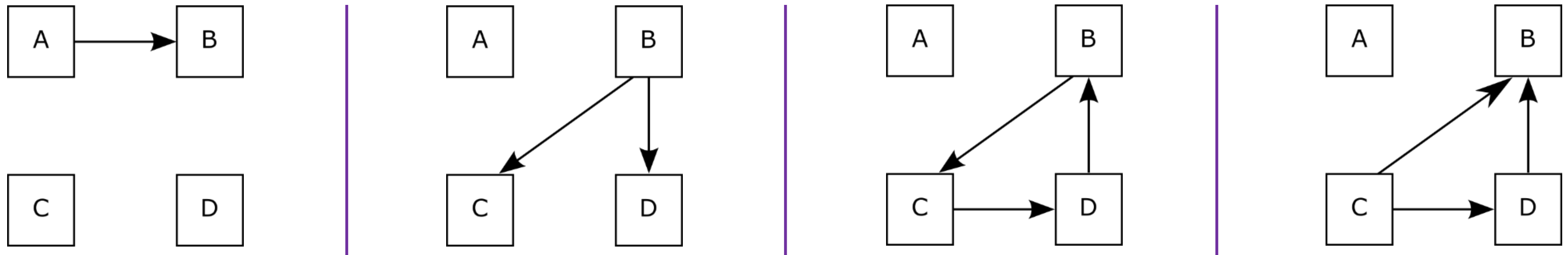
Separating volume,
topological and temporal factors

Three things that contribute to broadcast scores

- **Volume factors:**
  - the more messages a user sends, the higher their broadcast score is likely to be
- **Topological factors:**
  - a user who communicates with well-connected people is likely to have a higher broadcast score
- **Temporal factors:**
  - a user who sends a message at the start of a big flurry of messages is likely to have a higher broadcast score
  - this is right: it suggests the user has said something interesting/important that has sparked off a discussion or cascade of messaging

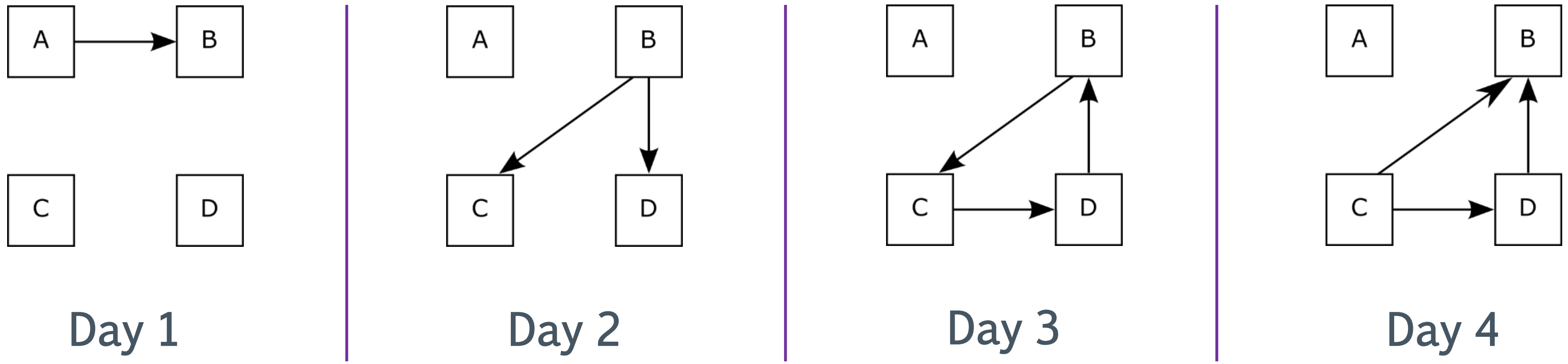
We want to untangle these factors to understand *why* each user's broadcast score is high or low

Example



Day 1 | Day 2 | Day 3 | Day 4

Example



- For alpha = 0.5 the broadcast scores are: A: 4.10, B: 6.18, C: 3.71, D: 2.93
- By eye we see that the broadcast score of node A is mainly due to the timing of edges
- Can we say something precise about this?

Idea: permutation testing

- Let us ask: **what happens to the broadcast scores when we randomly permute the timestamps of the edges?**

Day	Source	Target
1	A	B
2	B	C
2	B	D
3	B	C
3	C	D
3	D	B
4	C	B
4	C	D
4	D	B

Idea: permutation testing

› Let us ask: **what happens to the broadcast scores when we randomly permute the timestamps of the edges?**

Day	Source	Target
1	A	B
2	B	C
2	B	D
3	B	C
3	C	D
3	D	B
4	C	B
4	C	D
4	D	B

Day	Source	Target
3	A	B
4	B	C
1	B	D
2	B	C
2	C	D
3	D	B
3	C	B
4	C	D
4	D	B

Idea: permutation testing

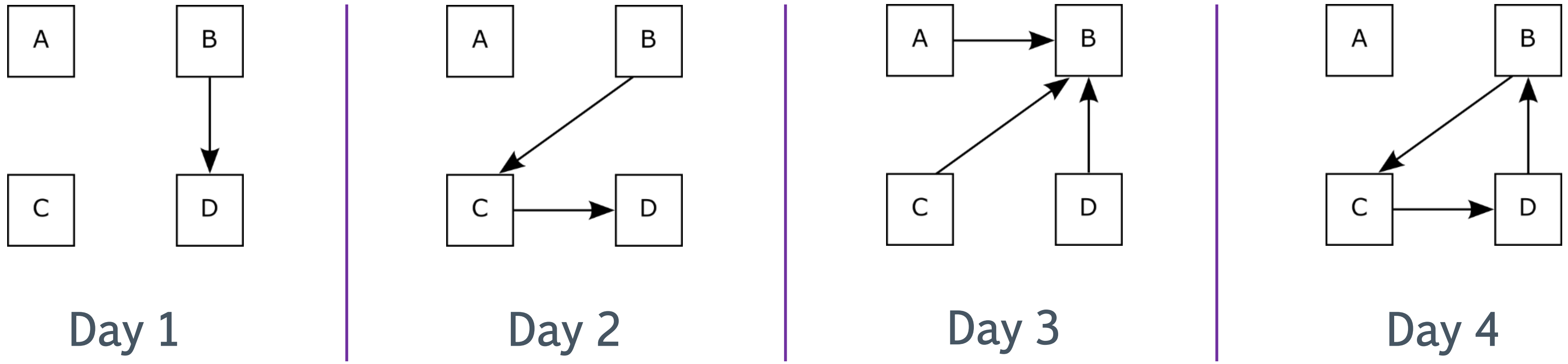
› Let us ask: **what happens to the broadcast scores when we randomly permute the timestamps of the edges?**

Day	Source	Target
1	A	B
2	B	C
2	B	D
3	B	C
3	C	D
3	D	B
4	C	B
4	C	D
4	D	B

Day	Source	Target
3	A	B
4	B	C
1	B	D
2	B	C
2	C	D
3	D	B
3	C	B
4	C	D
4	D	B

The timing of edges is scrambled, but the in- and out- degree of each node is preserved, as well as relationships (who sends to whom), and the number of edges on each day

In the permuted network…



- For alpha = 0.5 the broadcast scores are: A: 2.00, B: 5.75, C: 4.50, D: 3.00
- The broadcast score of A is much reduced: 2.00 down from 4.10.
- If we take the averages over all such permutations we get:

  A: 2.53, B: 5.25, C: 5.07, D: 4.13
- This confirms that the timing of edges was contributing a lot to A's broadcast score

Permutation testing

- So: **permuting the time column "undoes" the effect of the timing of edges**
  - but preserves who sent a message to whom, and how often, and the level of activity in the network each day
- **Permuting the "Target" column "undoes" the effect of the network topology**
  - but preserves the *number* of messages each user sent, and the *times* each user sent messages
- **Permuting any two columns** preserves only the number of edges sent and received by each user
  - so this **"undoes" the effects of both time and topology, leaving only the volume factor**

How to do this in practice

- The number of permutations grows extremely quickly, so we can't check them all for non-toy networks
- We could sample from the set of permutations.

How to do this in practice

- The number of permutations grows extremely quickly, so we can't check them all for non-toy networks
- We could sample from the set of permutations.
- But we can do better:
  - the main contribution of our paper is the derivation of closed formulae for the average broadcast scores over all permutations
  - so you don't need to look at any permutations at all

- We won't go into details here
  - involves jumping from a discrete to a continuous setting
  - solving appropriate matrix differential equations
  - which means doing algebra on matrix exponentials

Closed formulae for the mean broadcast scores

› When we permute two or three columns (the easiest case) the mean broadcast scores are:

$$\hat{\mathbf{b}}(t) = \mathbf{1}^T + \left(\frac{e^{\alpha[\mathbf{r}^T\mathbf{s}]t} - 1}{\mathbf{r}^T\mathbf{s}} \right) \mathbf{s}$$

$\hat{\mathbf{b}}(t)$ is the column vector of mean broadcast scores at time t

a vector of all 1s

$\mathbf{r}$ is the column vector of the average receiving rate of each vertex

$\mathbf{s}$ is the column vector of the average sending rate of each vertex

Summary

- presented dynamic communicability in evolving networks
  - a way of measuring how information propagates in an evolving network
  - generalises Katz centrality for (static) graphs
  - assigns broadcast scores to nodes, quantifying their "communication reach"

- showed results about broadcast scores and sentiment from a large Twitter dataset
  - people with the highest broadcast scores (comm. reach) use more positive sentiment
  - and less negative sentiment than the average user

- showed how to separate volume, temporal and topological factors which influence broadcast scores
  - initially using permutation testing
  - but closed formulae can be obtained