

# It's a matter of time!

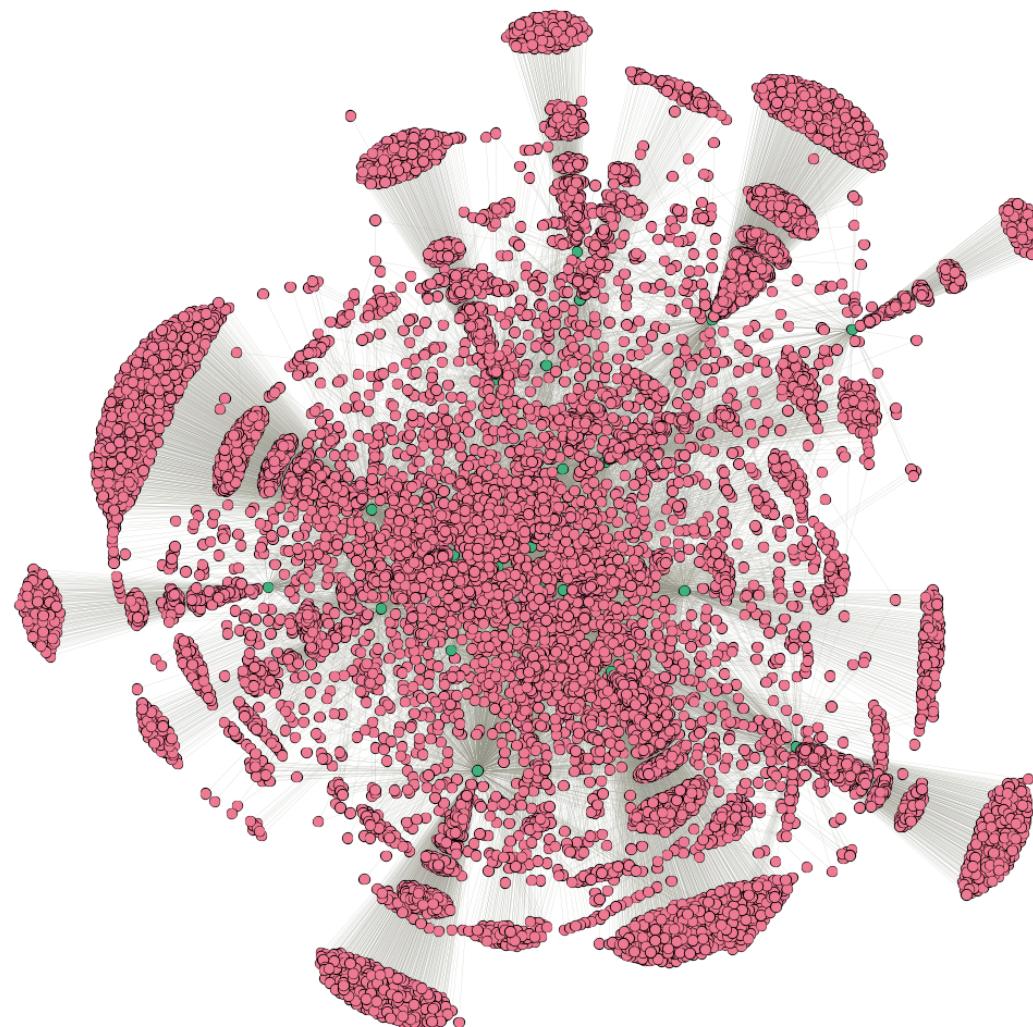
## Knowledge Discovery from Temporal Social Networks

**Fabíola S. F. Pereira (UFU)**  
João Gama (LIAAD, UPorto)  
Gina M. B. de Oliveira (UFU)

Short course – SBBD – Oct 2017

# Motivation

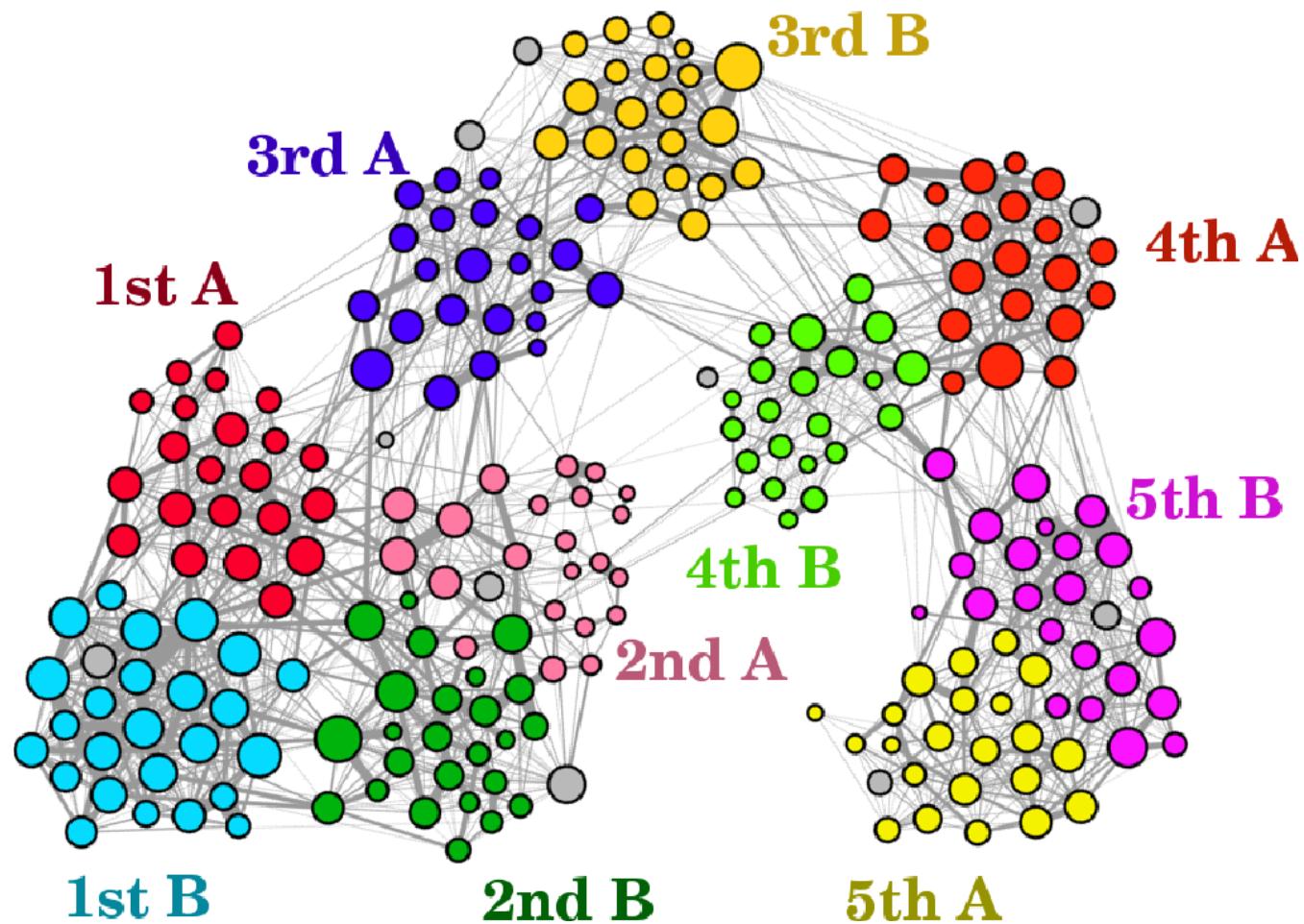
# Last.fm bipartite network (uSer-artist)



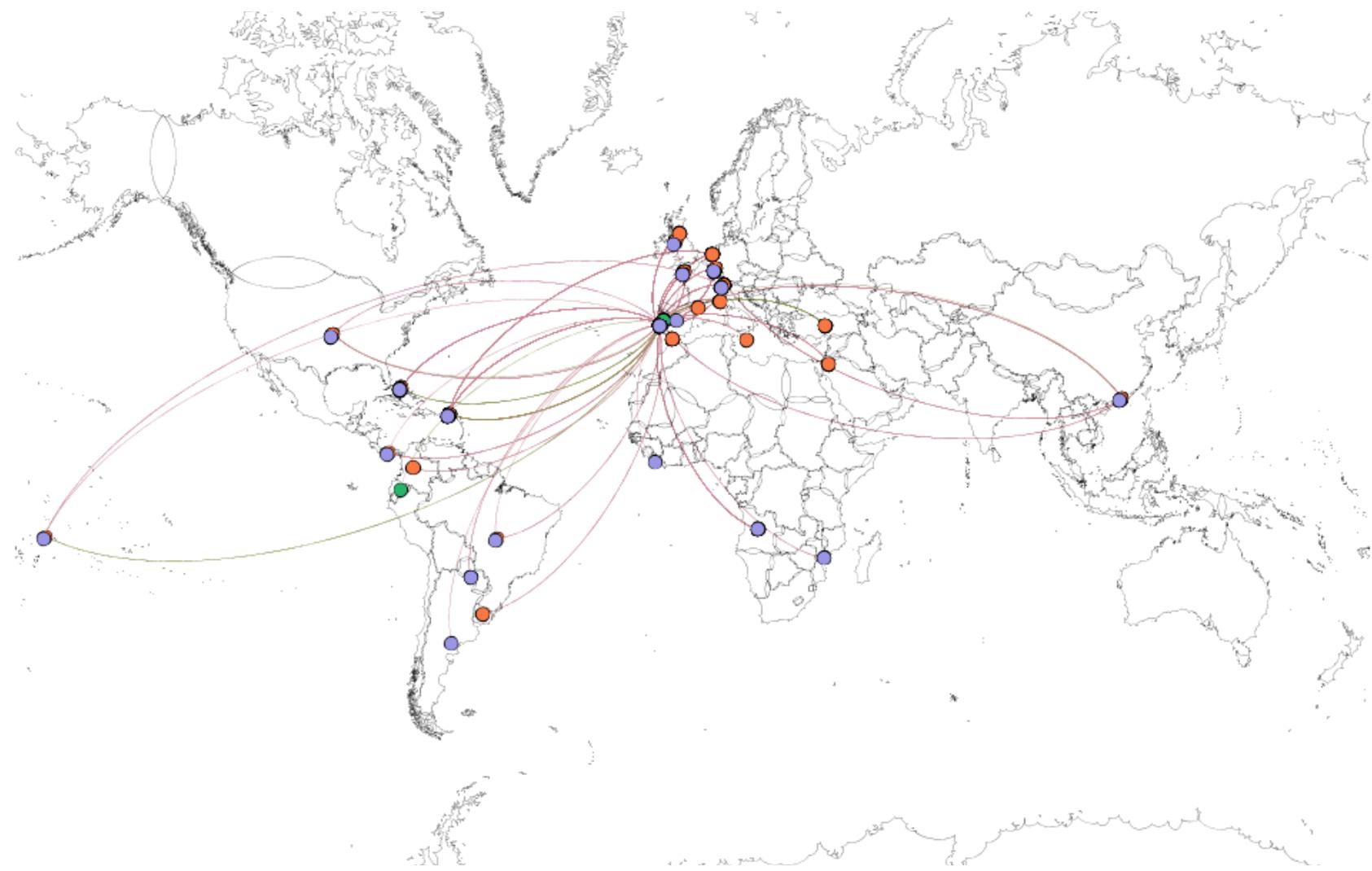
# Facebook Network



# Face-to-face contact in a primary school



# Panama papers - Portugal offShoreS



# Evolving Call network



# Outline

Overview

SNA  
Essentials

Temporal  
Social  
Networks

Processing  
Evolving  
Networks

Summary  
of  
Definitions

Case Study  
1: Twitter

Case Study  
2: Calls

Case Study  
3: Last.fm

Being a  
Data  
Scientist

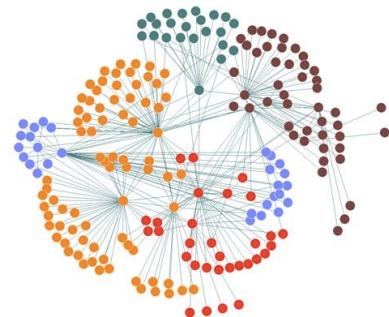
Final  
Remarks

# Overview

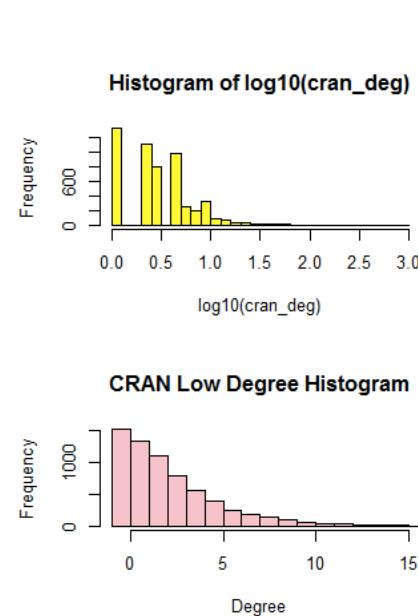
# Social Media Mining is the process of...



collecting



representing

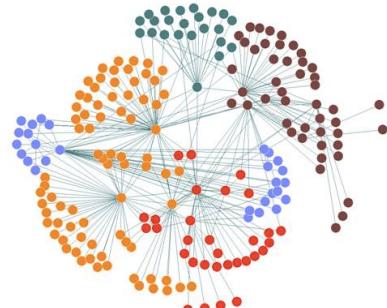


analyzing



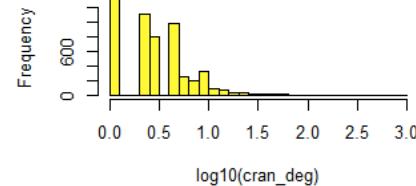
extracting  
patterns

# This short course is about...

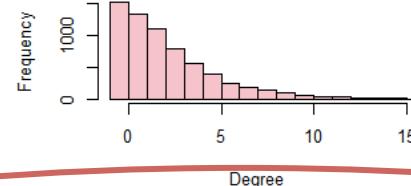


representing

Histogram of  $\log_{10}(\text{cran\_deg})$



CRAN Low Degree Histogram

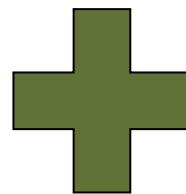


extracting  
patterns

over time

# Data Science for Social Media Mining

social  
theory



computational  
theory

**What we want to know about the vast social media world  
with computational tools?**

A Social media mining

data scientist

needs to do the

right questionS!

# **How Fast Will You Get a Response? Predicting Interval Time for Reciprocal Link Creation**

**Vachik S. Dave, Mohammad Al Hasan**  
Department of Computer & Information Science  
IUPUI, Indianapolis, USA  
[vsdave@iupui.edu](mailto:vsdave@iupui.edu), [alhasan@iupui.edu](mailto:alhasan@iupui.edu)

**Chandan K. Reddy**  
Department of Computer Science  
Virginia Tech, Arlington, USA  
[reddy@cs.vt.edu](mailto:reddy@cs.vt.edu)

**from ICWSM'17**

# **How Fast Will You Get a Response? Predicting Interval Time for Reciprocal Link Creation**

S. S. Dave, Mohammad Al Hasan  
Information Science

**Chandan K. Reddy**  
Department of Computer Science  
Virginia Tech, Arlington, USA  
[reddy@cs.vt.edu](mailto:reddy@cs.vt.edu)

# **What Comments Did I Get? How Post and Comment Characteristics Predict Interaction Satisfaction on Facebook**

Shruti Sannon, Yoon Hyung Choi, Jessie G. Taft, Natalya N. Bazarova

Department of Communication, Cornell University  
[{ss3464, yc863, jgt43, nnb8}@cornell.edu](mailto:{ss3464, yc863, jgt43, nnb8}@cornell.edu)

**from ICWSM'17**

# How Fast Will You Get a Response? Predicting Interval Time for Reciprocal Link Creation

S. Dave, Mohammad Al Hasan  
Information Science

Chandan K. Reddy  
Department of Computer Science  
Virginia Tech, Arlington, USA  
reddy@cs.vt.edu

# What Comments Did I Get? How Post and Comment Characteristics Predict Interaction Satisfaction

Shruti Sannon, Yoon H.

ook

Bazarova

# Nasty, Brutish, and Short: What Makes Election News Popular on Twitter?

Sophie Chou  
MIT Media Lab  
Cambridge, MA, USA  
soph@media.mit.edu

Deb Roy  
MIT Media Lab  
Cambridge, MA, USA  
dkroy@media.mit.edu

from ICWSM'17

# How Fast Will You Get a Response? Predicting Interval Time for Reciprocal Link Creation

**Are There Gender Differences in Professional Self-Promotion? Characteristics of Recent MBA Graduates**

**Chandan K. Reddy**  
Department of Computer Science  
Virginia Tech, Arlington, USA  
reddy@cs.vt.edu

**Rajlakshmi De**  
North Carolina State University (USA)  
rde2@ncsu.edu

**Kaylyn Frazier**  
Google Inc. (USA)  
kfrazier@google.com

**Jim Hamilton**  
LinkedIn (USA)  
jihamilton@linkedin.com

**Sophie Chou**  
MIT Media Lab  
Cambridge, MA, USA  
soph@media.mit.edu

**Nikolai Avteniev**  
LinkedIn (USA)  
navteniev@linkedin.com

**Kristen M. Altenburger**  
Stanford University (USA)  
kaltenb@stanford.edu

**Are There Gender Differences in Professional Self-Promotion? Characteristics of Recent MBA Graduates**

**Chandan K. Reddy**  
Department of Computer Science  
Virginia Tech, Arlington, USA  
reddy@cs.vt.edu

**Rajlakshmi De**  
North Carolina State University (USA)  
rde2@ncsu.edu

**Kaylyn Frazier**  
Google Inc. (USA)  
kfrazier@google.com

**Jim Hamilton**  
LinkedIn (USA)  
jihamilton@linkedin.com

**Sophie Chou**  
MIT Media Lab  
Cambridge, MA, USA  
soph@media.mit.edu

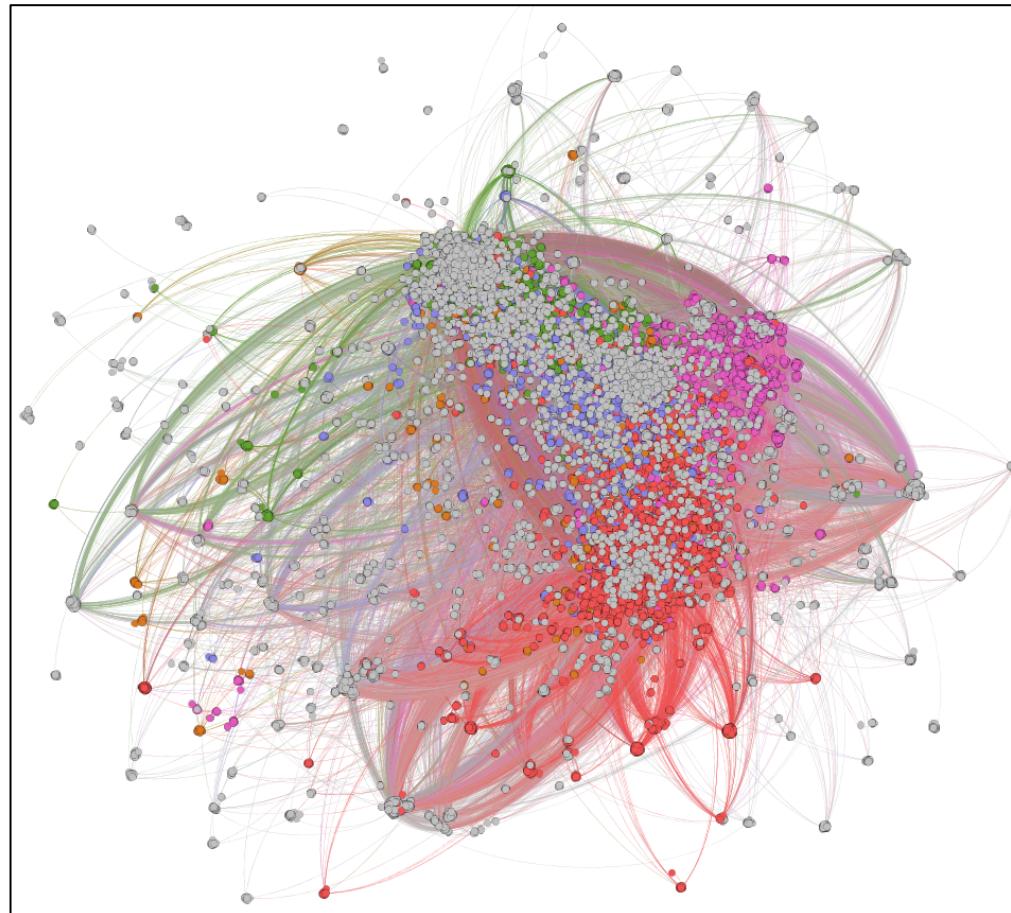
**Nikolai Avteniev**  
LinkedIn (USA)  
navteniev@linkedin.com

**Kristen M. Altenburger**  
Stanford University (USA)  
kaltenb@stanford.edu

## What Makes Electrical Engineers Tick?

from ICWSM'17

# The network structure is our



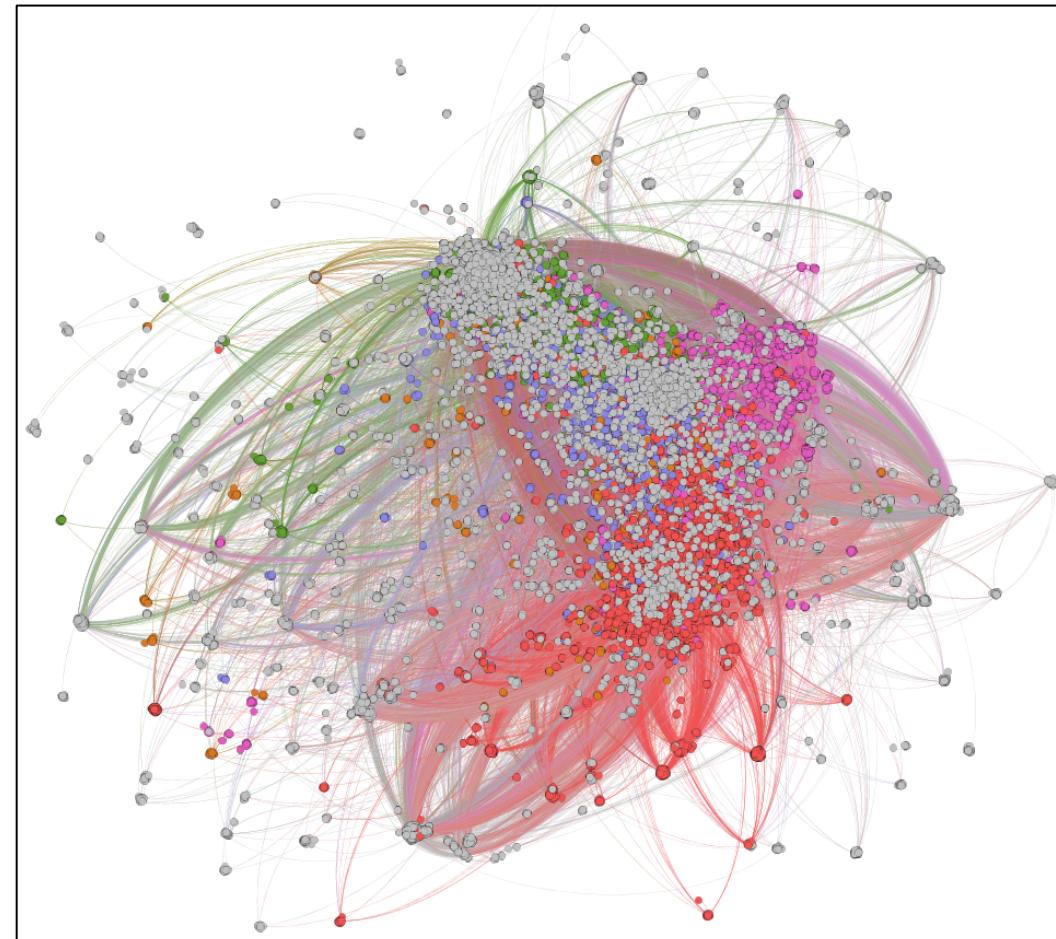
## Source of knowledge

# Social Network Analysis

*How can we **measure** the **influence** of individuals in a social network?*

*How can we **analyze** the **behavior** of individuals?*

*Who are the **most important people** in a social network?*



[Epinions Ego Trust Network from Massa & Avesani 2007 dataset]

# Social Network Analysis

*Sociology*

*Algorithms*

*Data Mining*

*over **network** structure*

# Social Network Analysis

*Social Network Analysis*

$\neq$

*Social Media*

# What are Social Networks?

- ✓ Real life (explicit)
- ✓ Online Social Networks (OSN) (explicit)
- ✓ Derived (implicit)

- |                           |                             |
|---------------------------|-----------------------------|
| ✓ e-mail networks,        | ✓ trust networks,           |
| ✓ citation networks,      | ✓ contact networks,         |
| ✓ co-author networks,     | ✓ music listening networks, |
| ✓ terrorist networks,     | ✓ sexual networks,          |
| ✓ telecom calls networks, | ✓ organized crime networks  |

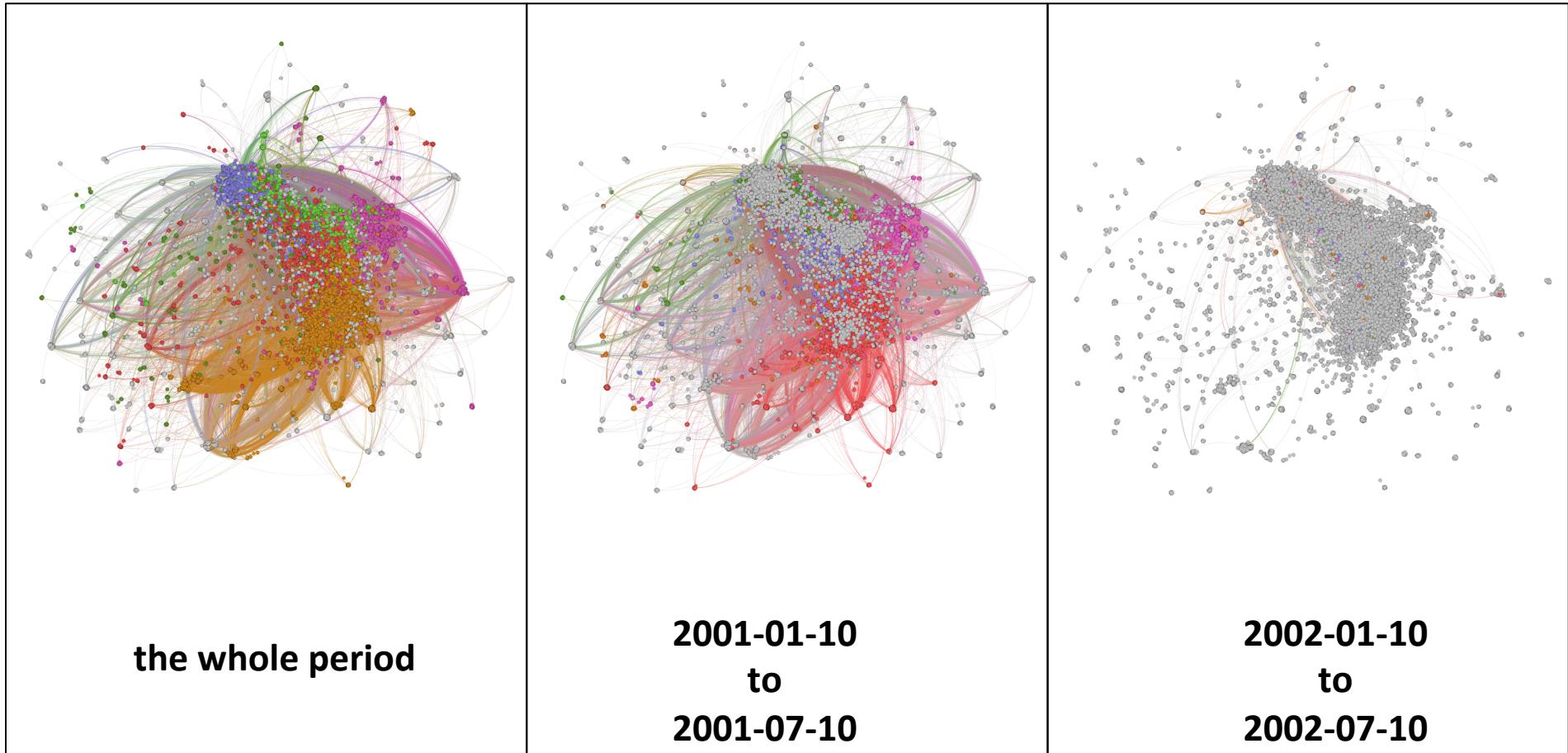
and whatever social interaction represented as network!

But wait....

Are Social networks  
evolving  
as time flies?

Or are they static?

# Evolving the Epinions Ego Trust Network



**top-6 modularity classes**

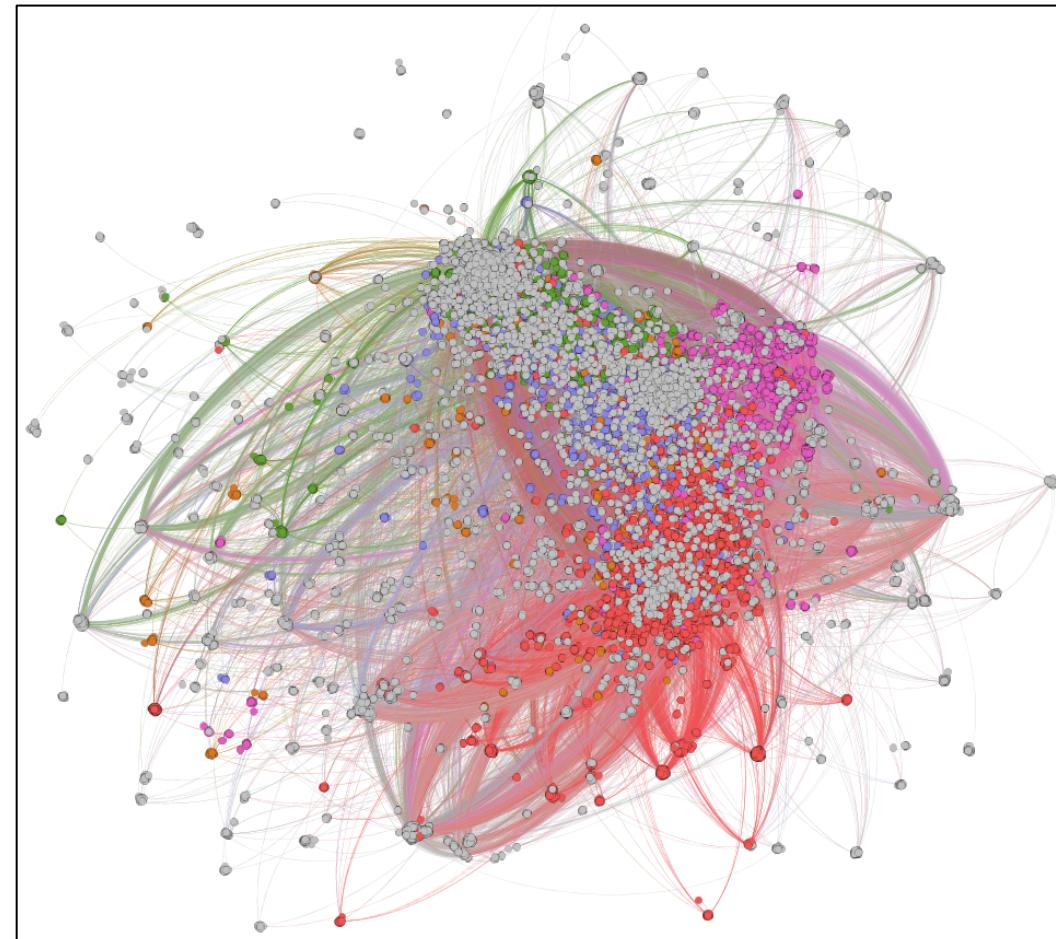
**It's a matter of time!**

# Social Network Analysis

*How can we **measure** the **influence** of individuals in a social network?*

*How can we **analyze** the **behavior** of individuals?*

*Who are the **most important people** in a social network?*



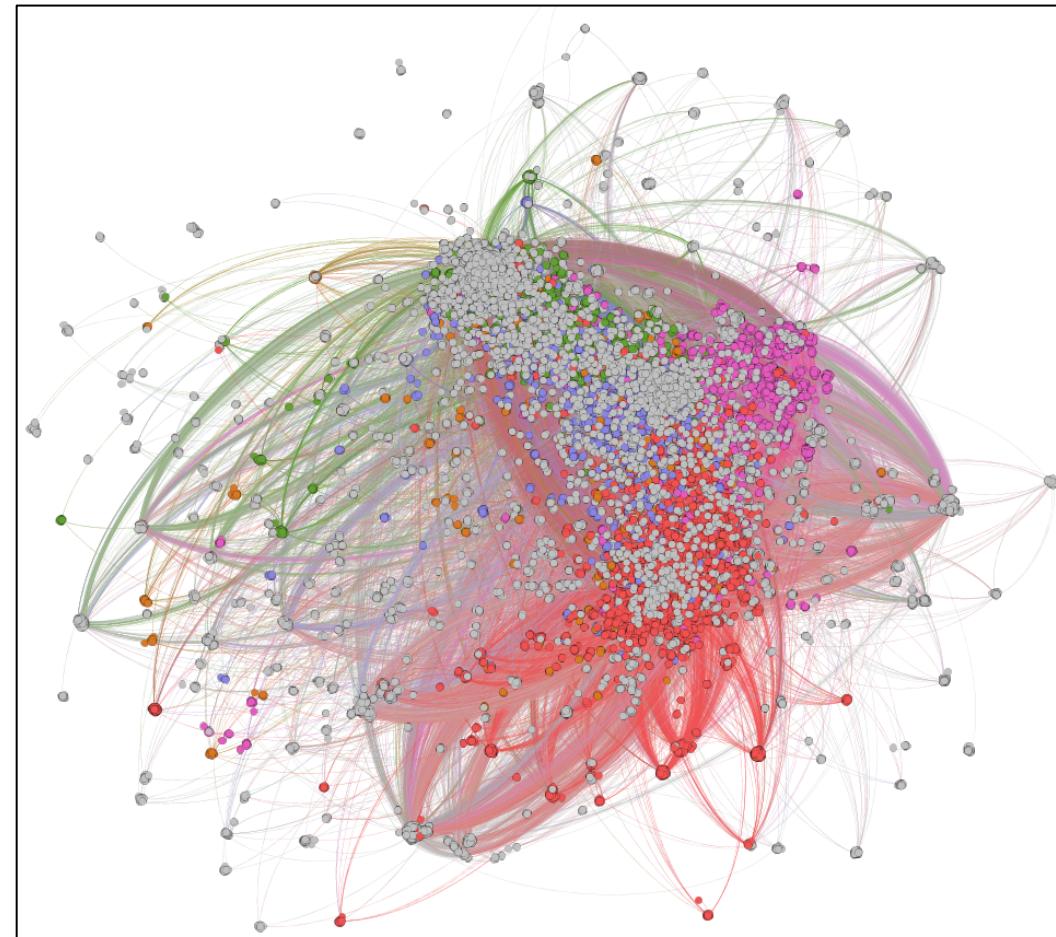
[Epinions Ego Trust Network from Massa & Avesani 2007 dataset]

# Dynamic Social Network Analysis

*How can we **measure** the **influence** of individuals in a social network?*

*How can we **analyze** the **behavior** of individuals?*

*Who are the **most important people** in a social network?*



[Epinions Ego Trust Network from Massa & Avesani 2007 dataset]

# Dynamic Social Network Analysis

*How can we **measure**  
**the influence** of  
individuals in a social  
network?*

*How can we **analyze the**  
**behavior** of individuals?*

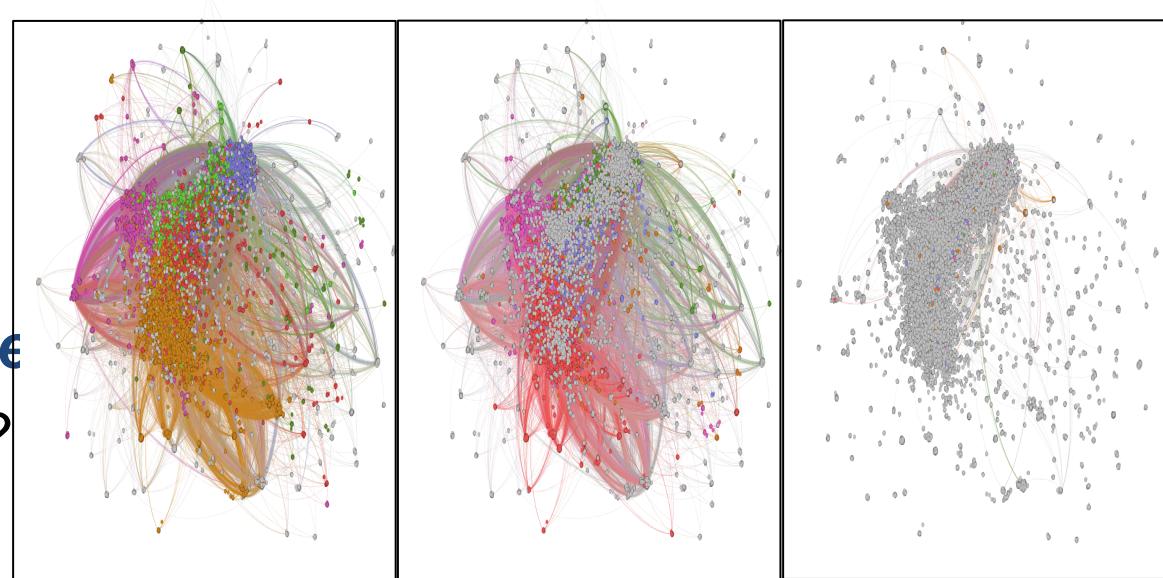
*Who are the **most**  
**important people** in a  
social network?*

# Dynamic Social Network Analysis

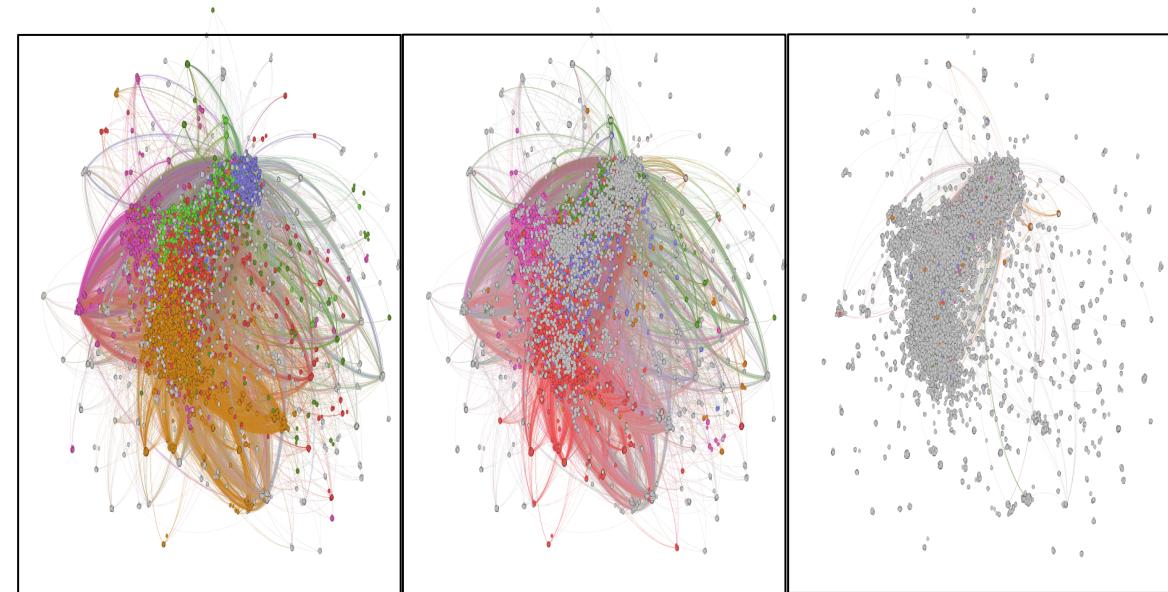
*How can we **measure**  
**the influence** of  
individuals in a social  
network?*

*How can we **analyze** the  
**behavior** of individuals?*

*Who are the **most**  
**important people** in a  
social network?*

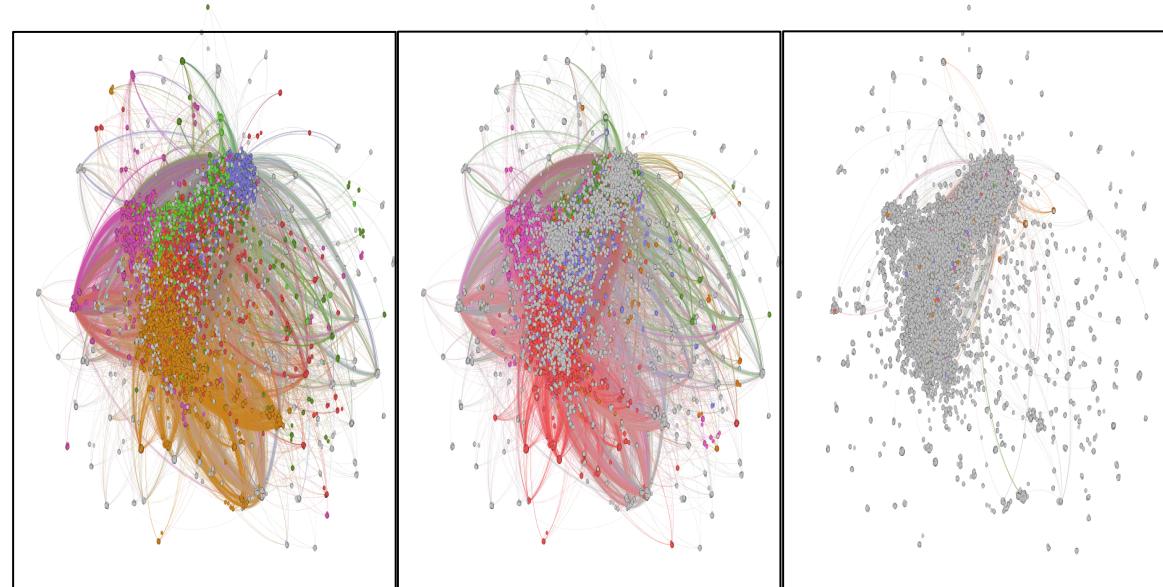


# Dynamic Social Network Analysis



# Dynamic Social Network Analysis

Who are likely to  
**become** friends next?

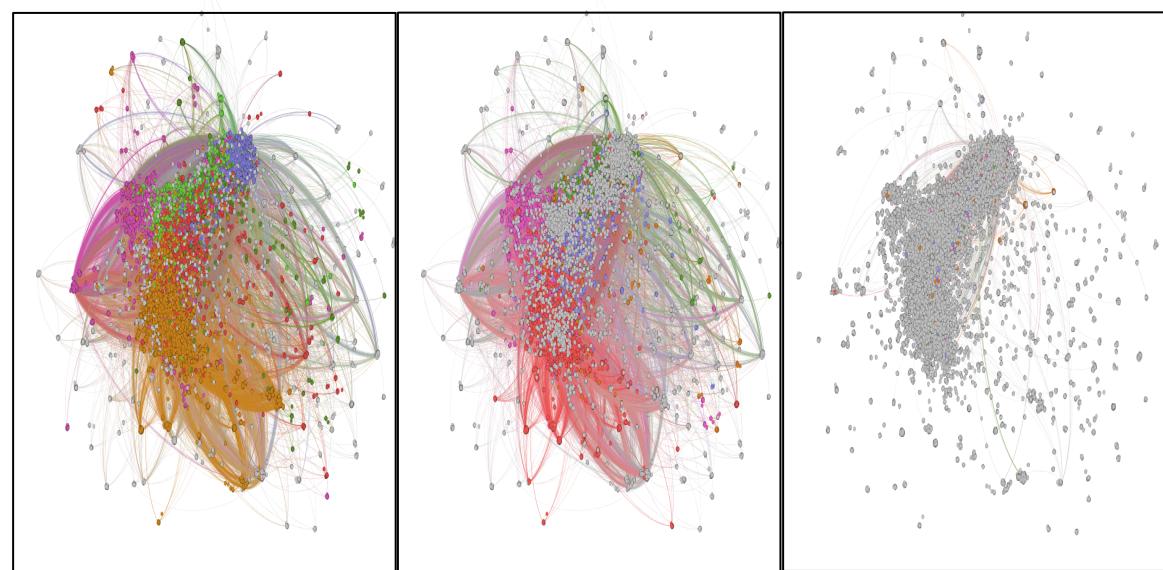


# Dynamic Social Network Analysis

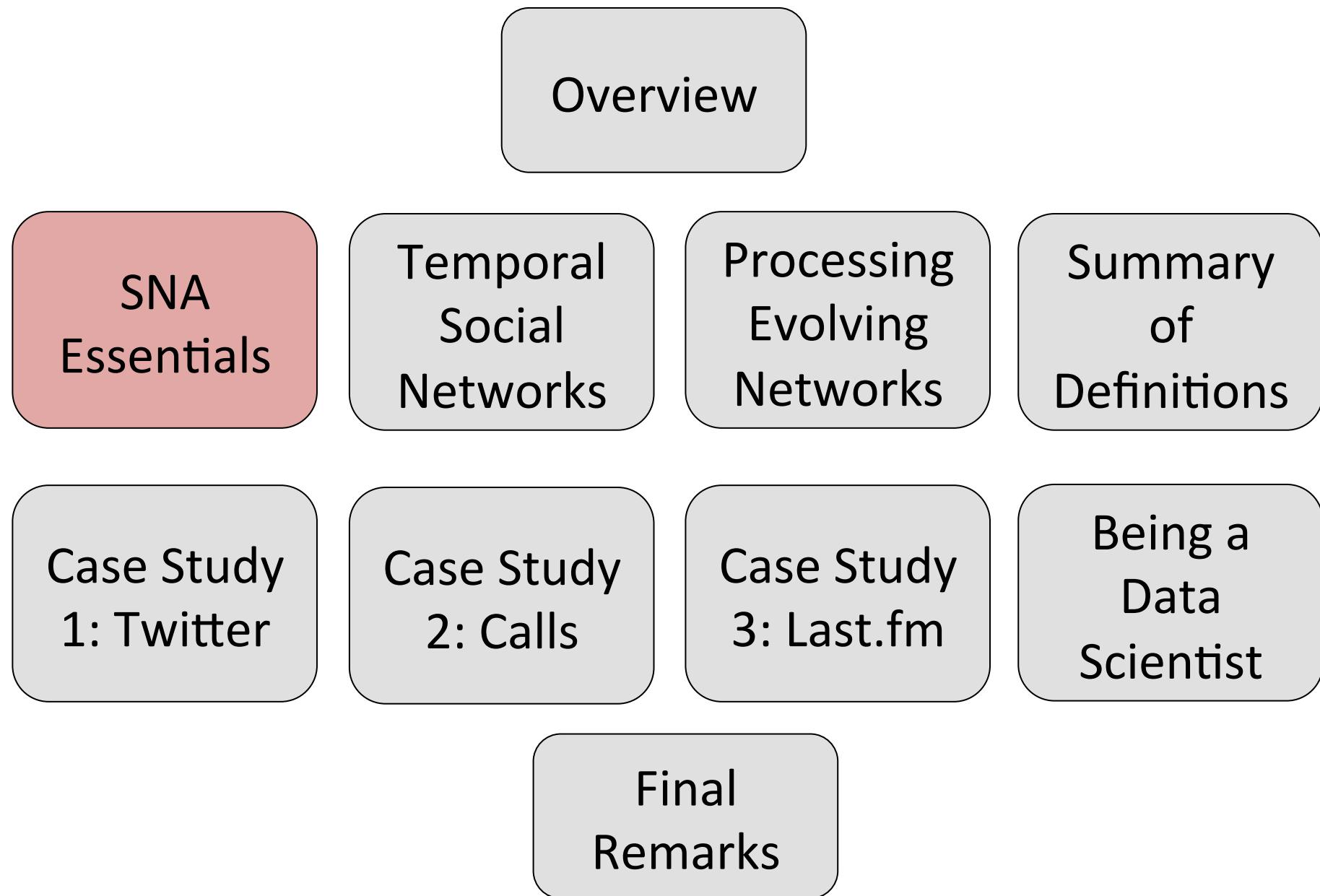
Who are likely to  
**become** friends next?

How does the **social**  
**network evolve** over  
time?

Can I **process** the  
whole network  
indefinitely?

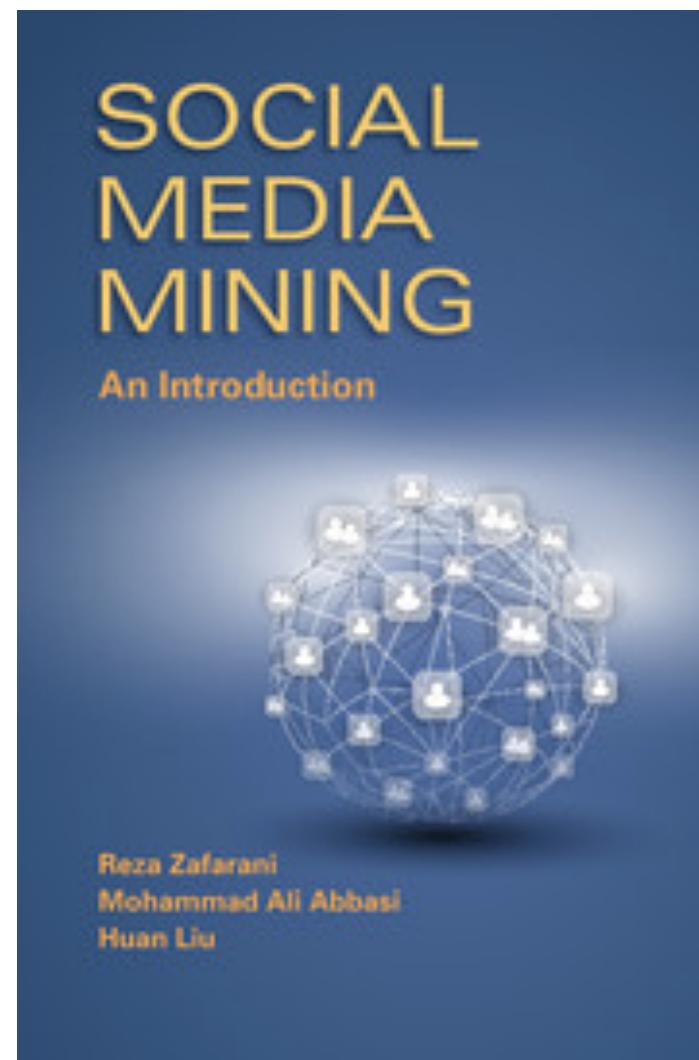


# Outline



# SNA Essentials

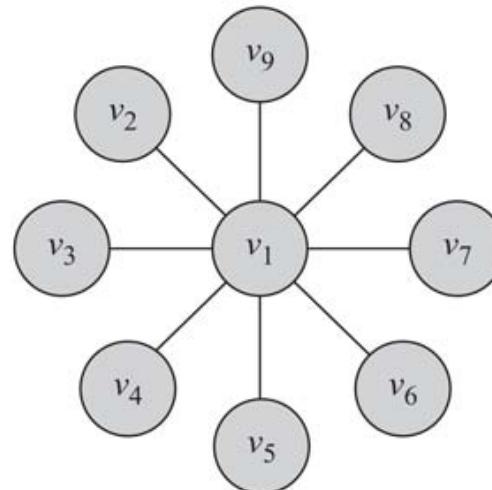
# SNA Essentials



# Network Measures

## Degree Centrality

$C_d(v_i) = d_i \rightarrow$  number of adjacent edges



$$C_d(v_1) = d_1 = 8$$

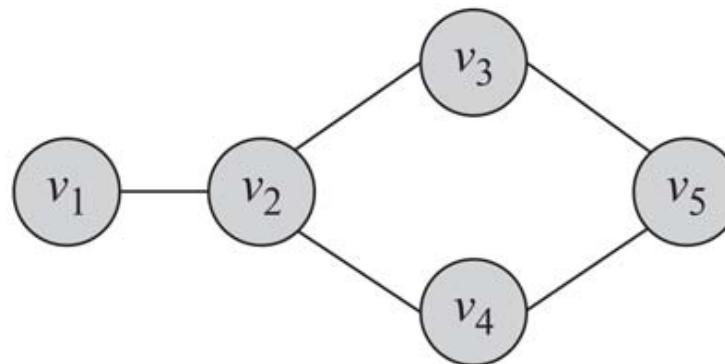
# Network Measures

## Betweenness Centrality

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

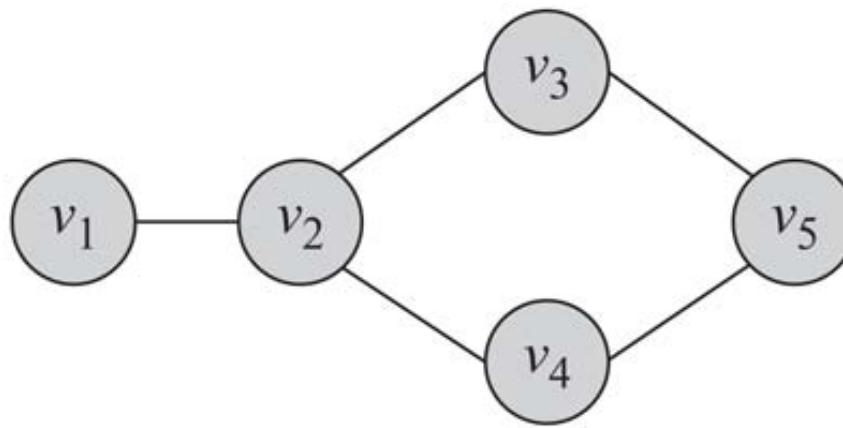
number of shortest paths  
from s to t passing  
through  $v_i$

number of shortest  
paths from s to t



# Network Measures

## Betweenness Centrality



$$C_b(v_2) = 2 \times \left( \underbrace{(1/1)}_{s=v_1, t=v_3} + \underbrace{(1/1)}_{s=v_1, t=v_4} + \underbrace{(2/2)}_{s=v_1, t=v_5} + \underbrace{(1/2)}_{s=v_3, t=v_4} + \underbrace{0}_{s=v_3, t=v_5} + \underbrace{0}_{s=v_4, t=v_5} \right)$$
$$= 2 \times 3.5 = 7$$

# Network Measures

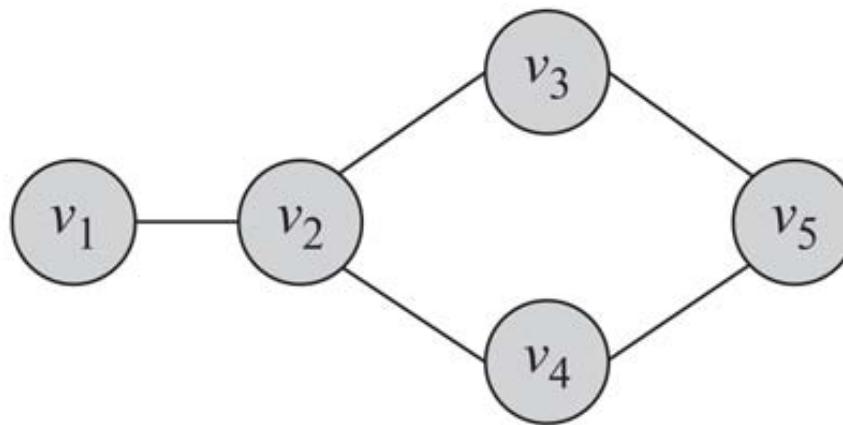
## Closeness Centrality

$$C_c(v_i) = \frac{1}{\bar{l}_{v_i}}$$

→ vi's average shortest path length to other nodes

# Network Measures

## Closeness Centrality



$$C_c(v_1) = 1 / ((1 + 2 + 3 + 4)/4) = 0.5,$$

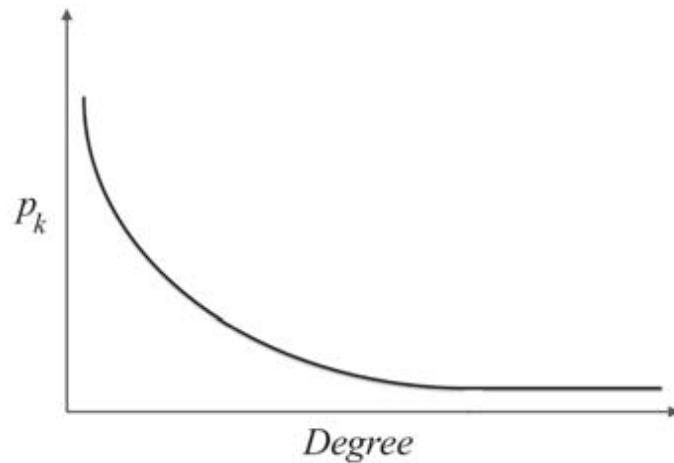
$$C_c(v_2) = 1 / ((1 + 1 + 1 + 2)/4) = 0.8, \quad \leftarrow$$

$$C_c(v_3) = C_b(v_4) = 1 / ((1 + 1 + 2 + 2)/4) = 0.66,$$

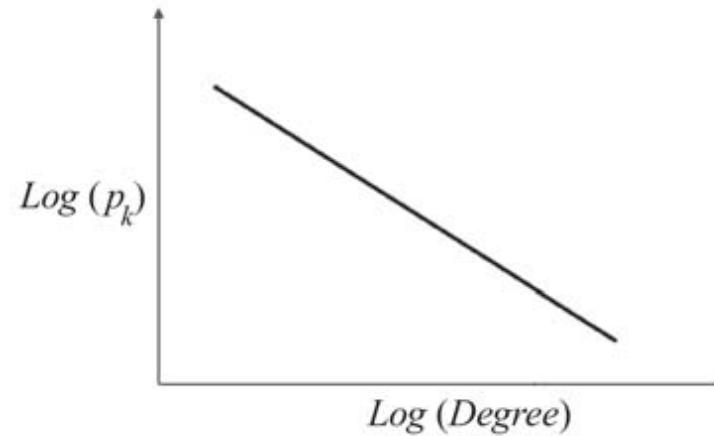
$$C_c(v_5) = 1 / ((1 + 1 + 2 + 3)/4) = 0.57.$$

# Network Models

## Properties of Real-World Networks



(a) Power-Law Degree Distribution



(b) Log-Log Plot of Power-Law Degree Distribution

Degree distribution - Power Law

# Network Models

## Properties of Real-World Networks

Average Local Clustering Coefficient in Real-World Networks

Web	Facebook	Flickr	LiveJournal	Orkut	YouTube
0.081	0.14 (with 100 friends)	0.31	0.33	0.17	0.13

Clustering Coefficient - High average

# Network Models

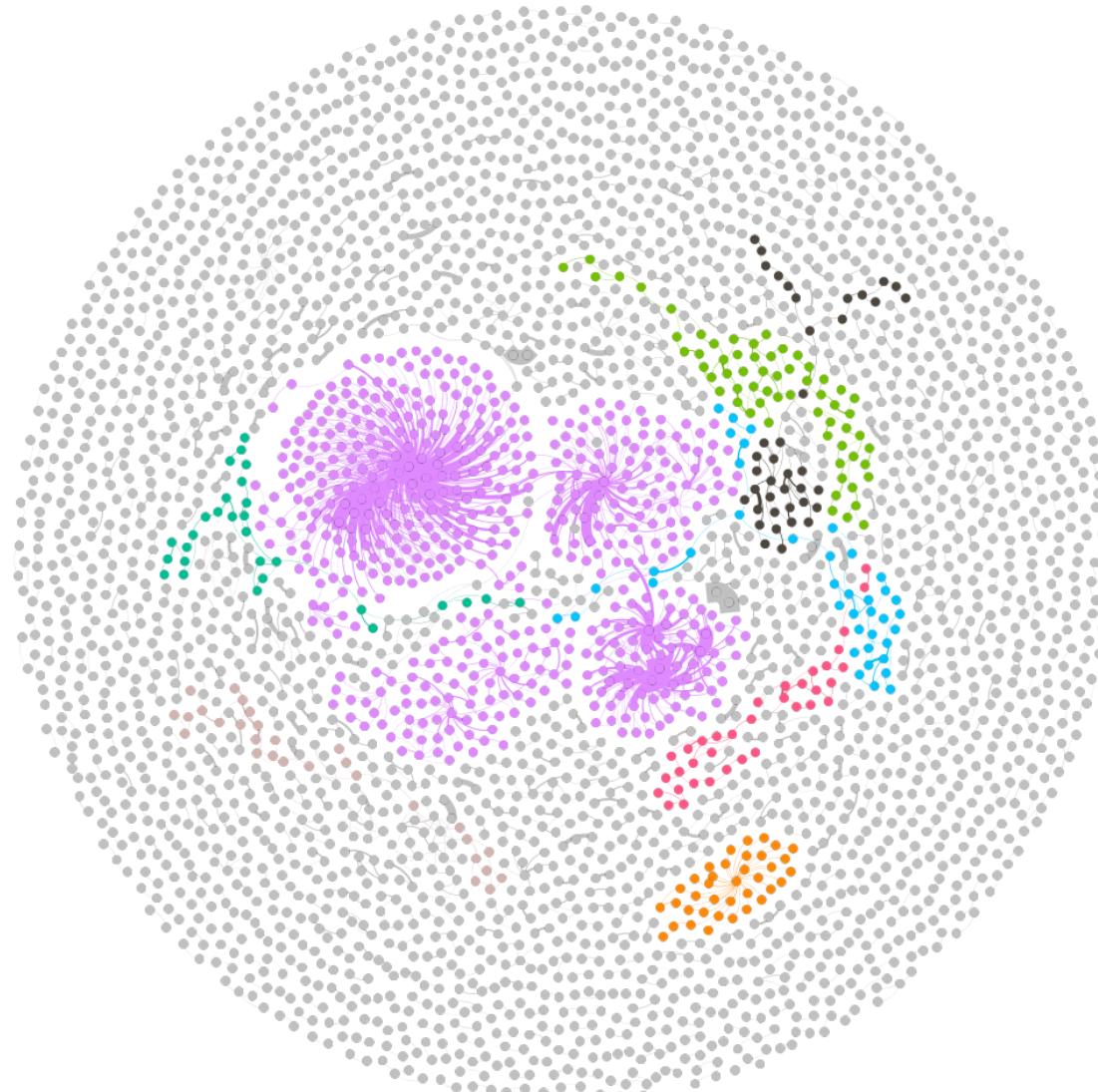
## Properties of Real-World Networks

Average Path Length in Real-World Networks

Web	Facebook	Flickr	LiveJournal	Orkut	YouTube
16.12	4.7	5.67	5.88	4.25	5.10

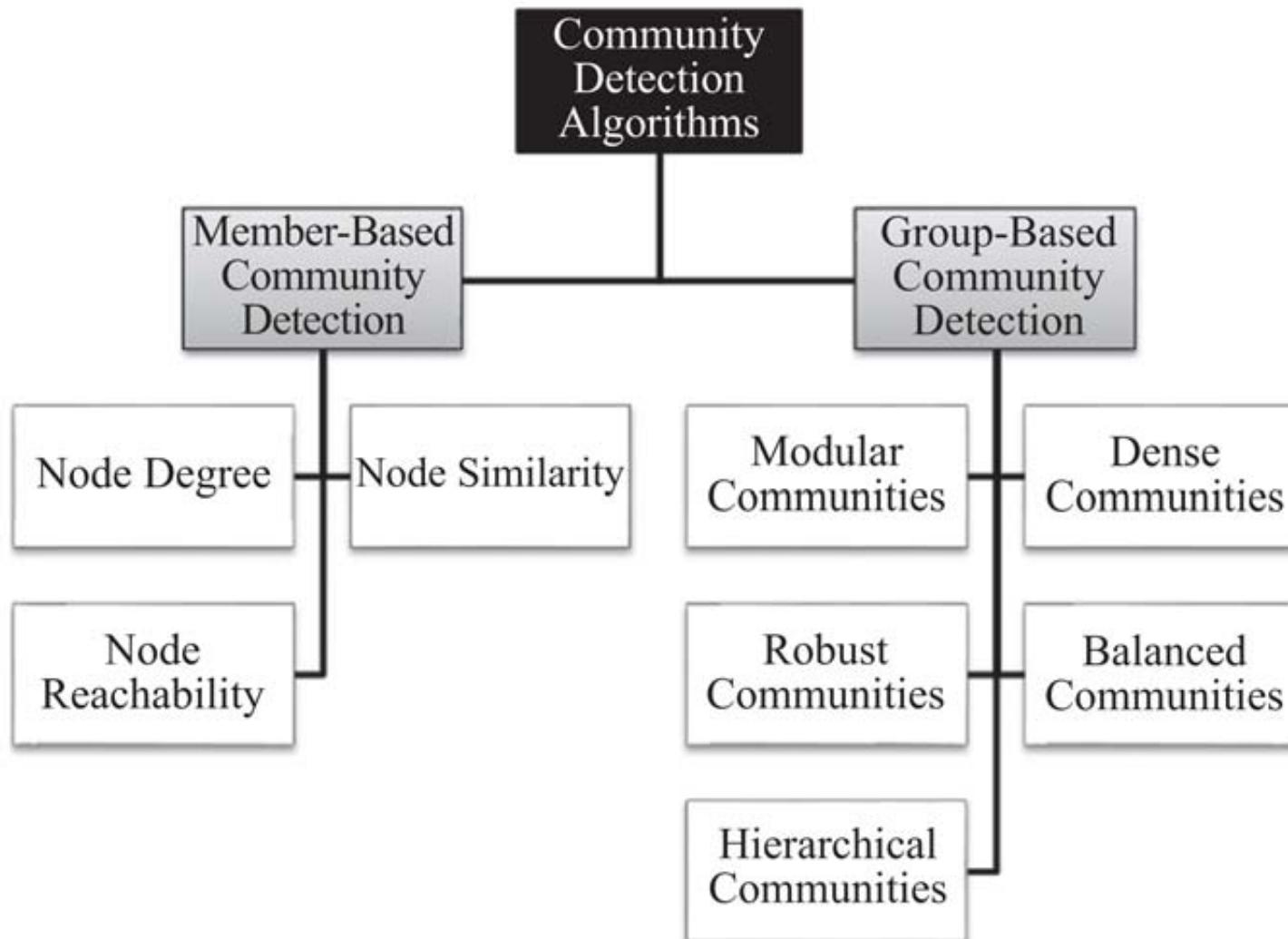
Average Path Length - Small World

# Communities



Telecom calls network

# Communities



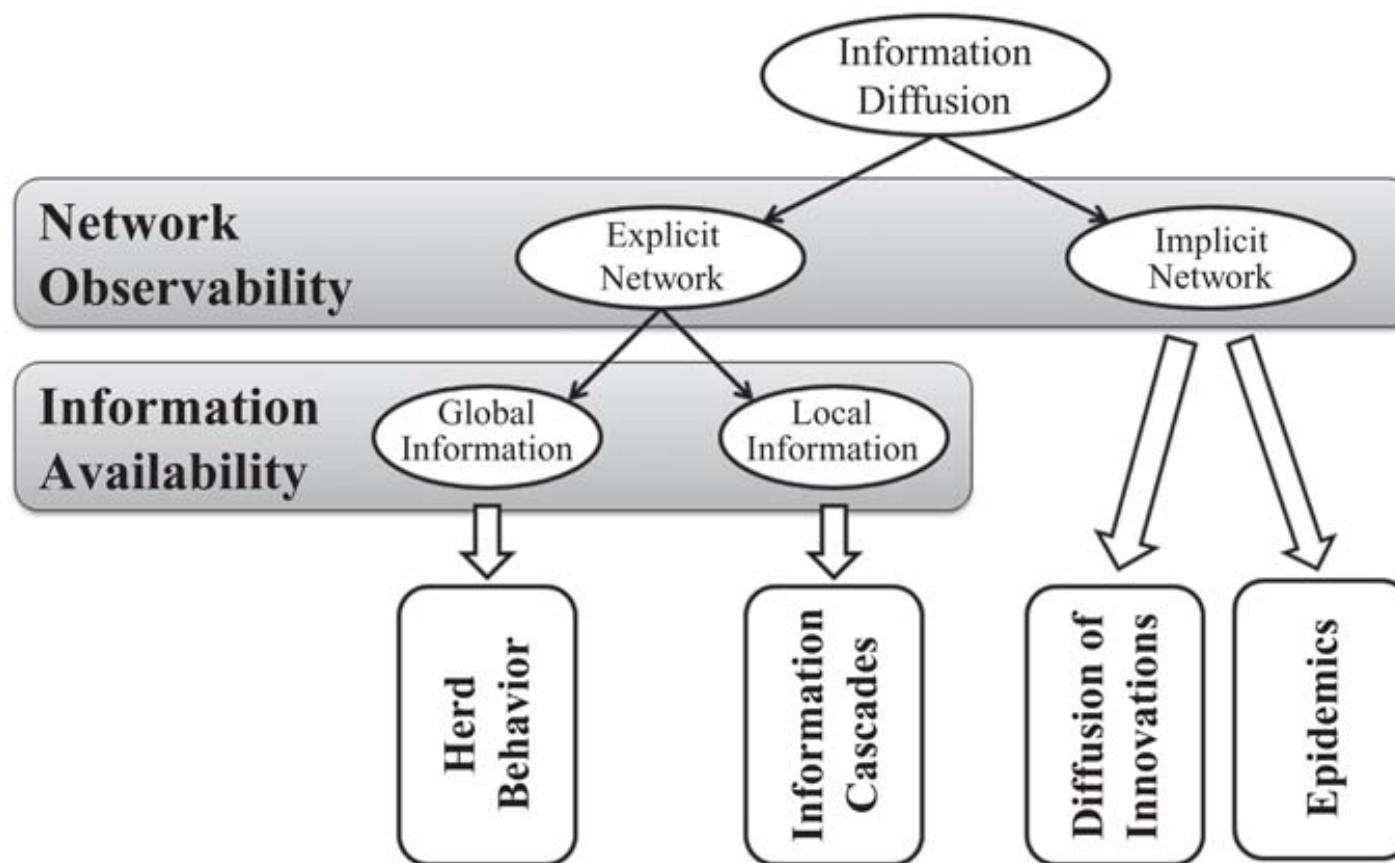
# Information Diffusion

*The process by which a piece of information  
(knowledge) is spread and reaches  
individuals through interactions*

# Information Diffusion

*The process by which a piece of information  
(knowledge) is spread and reaches  
individuals through interactions*

# Information Diffusion

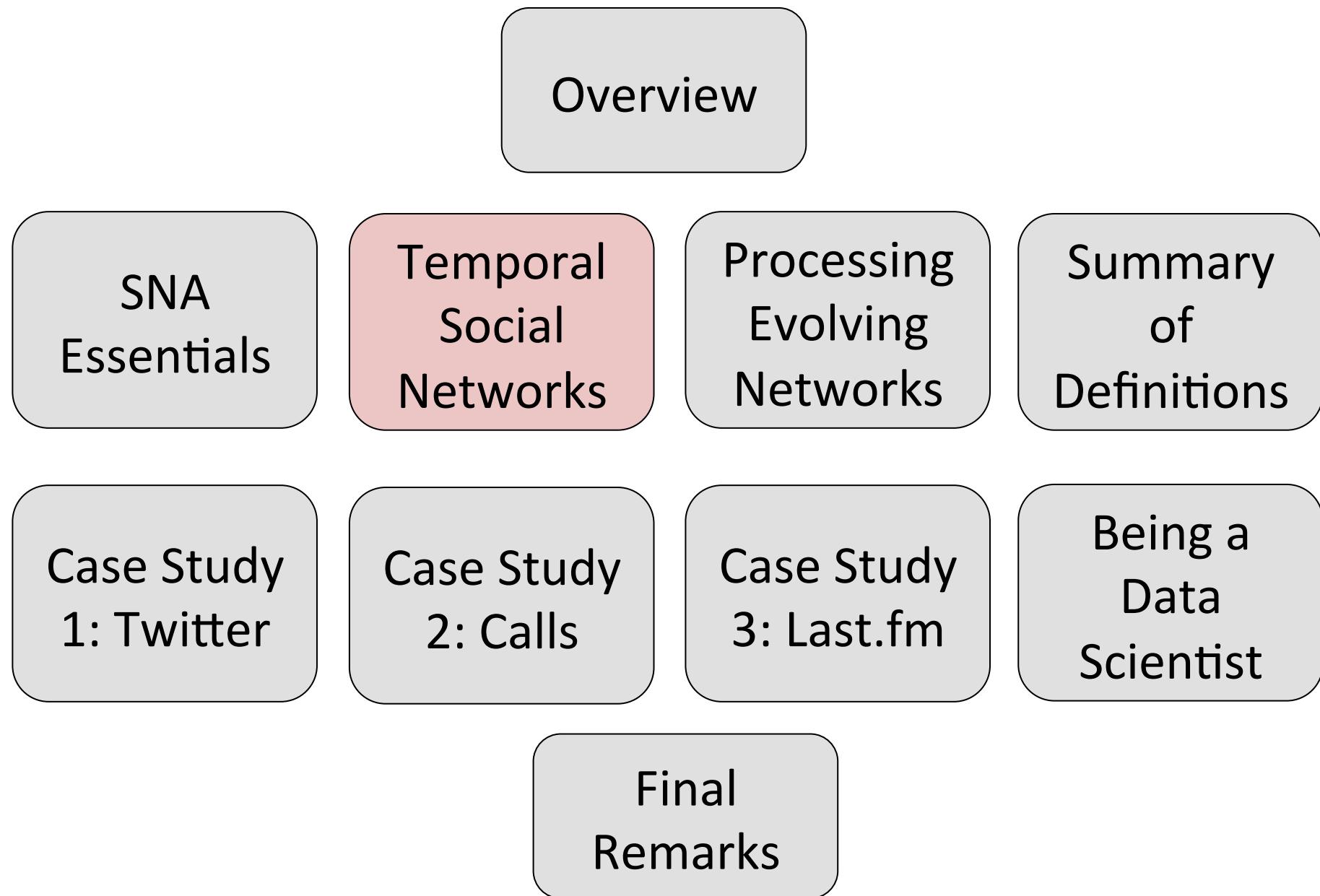


# HANDS ON!!

[www.lsi.facom.ufu.br/~fabiola/sbbd2017](http://www.lsi.facom.ufu.br/~fabiola/sbbd2017)

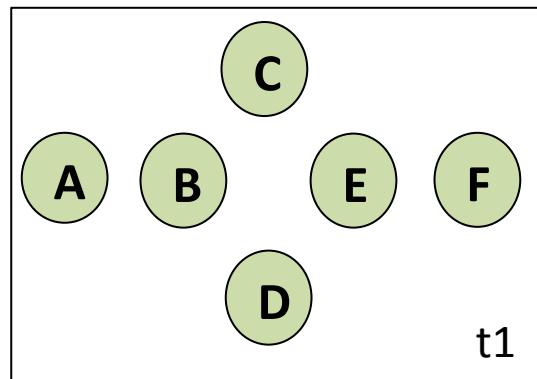


# Outline

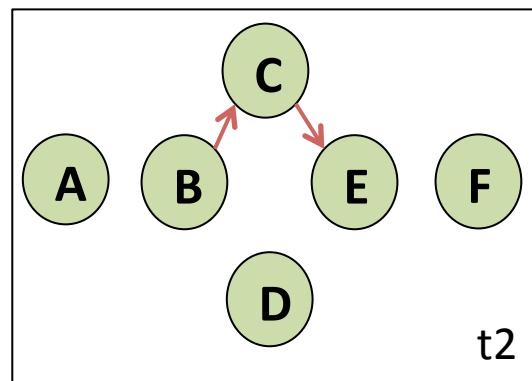


# Temporal Social Networks

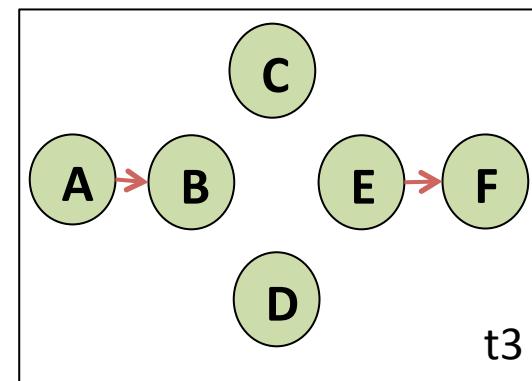
# Temporal Networks



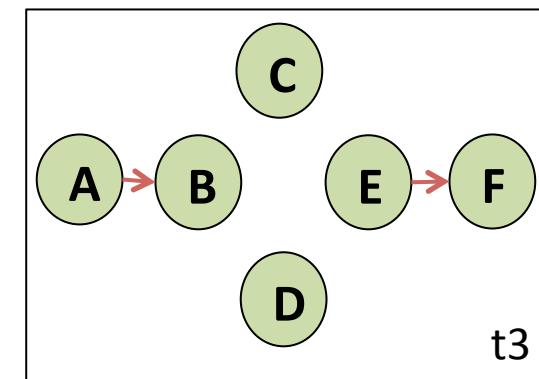
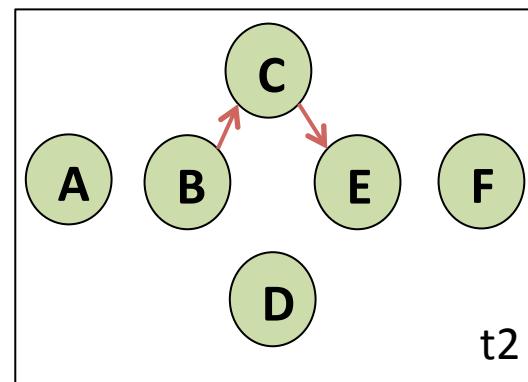
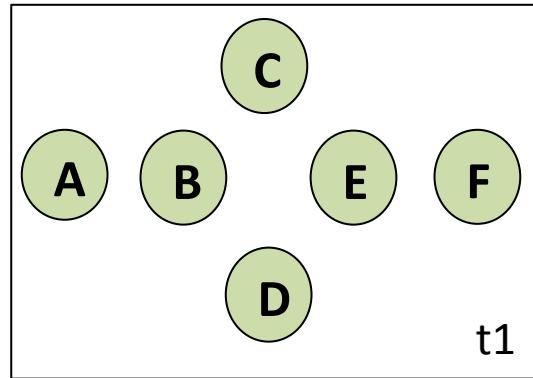
# Temporal Networks

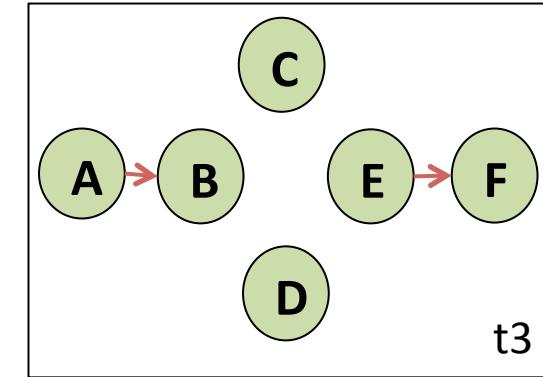
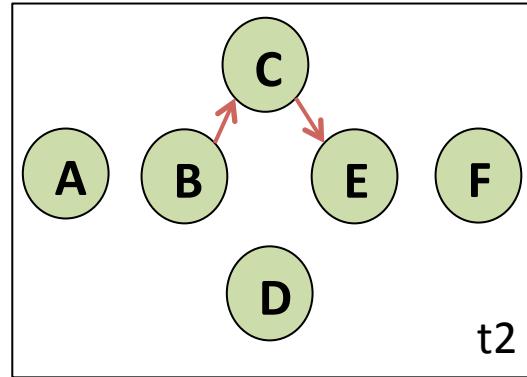
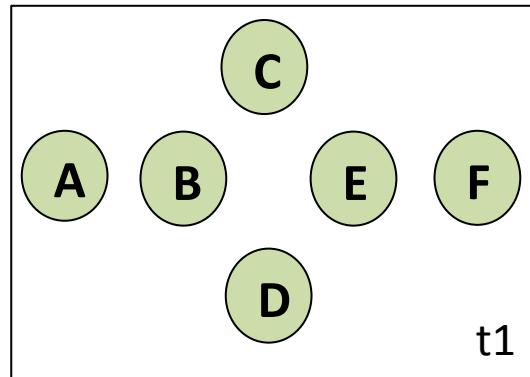


# Temporal Networks

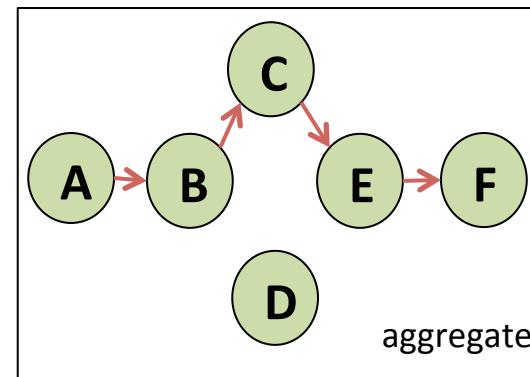


# Temporal Networks

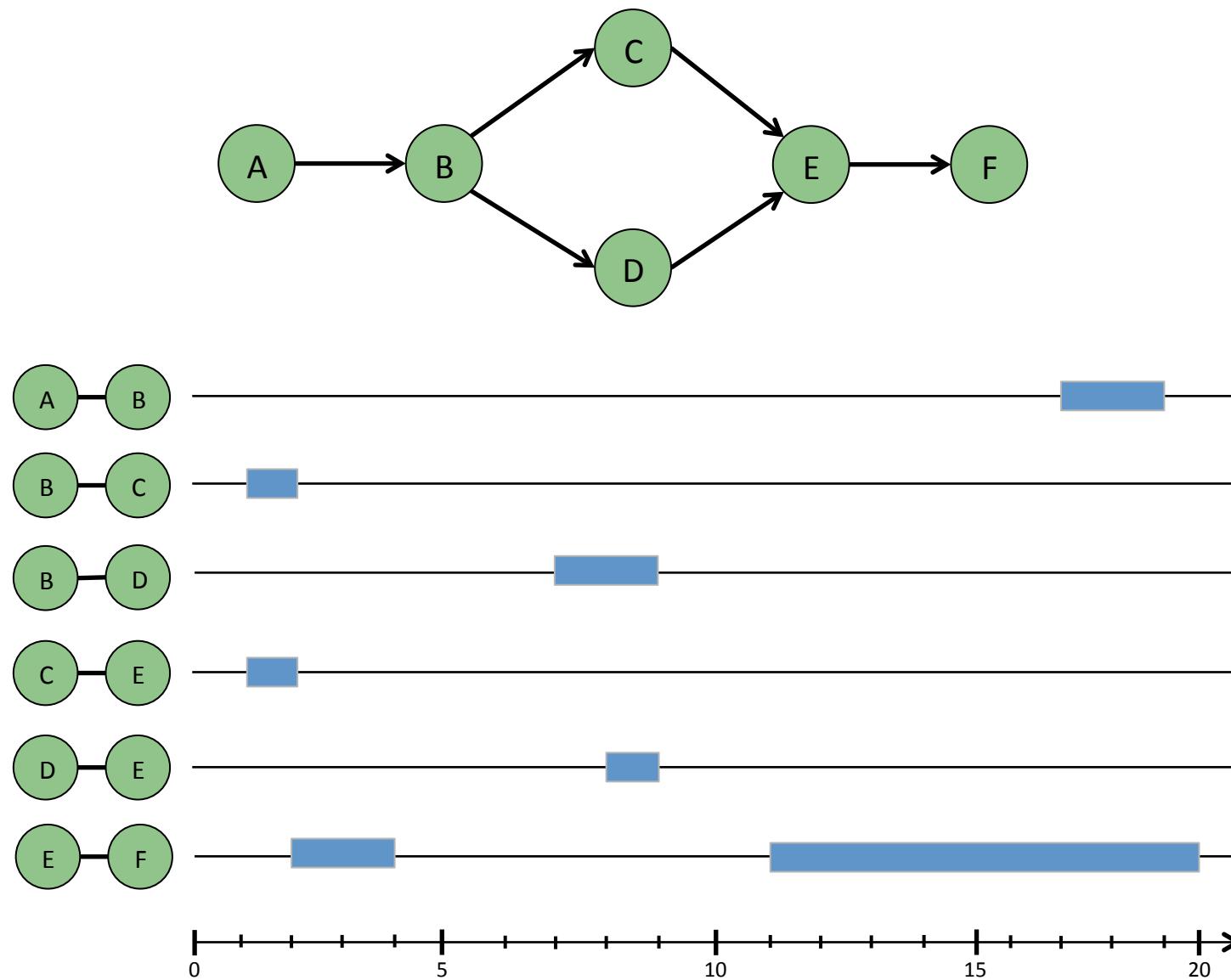




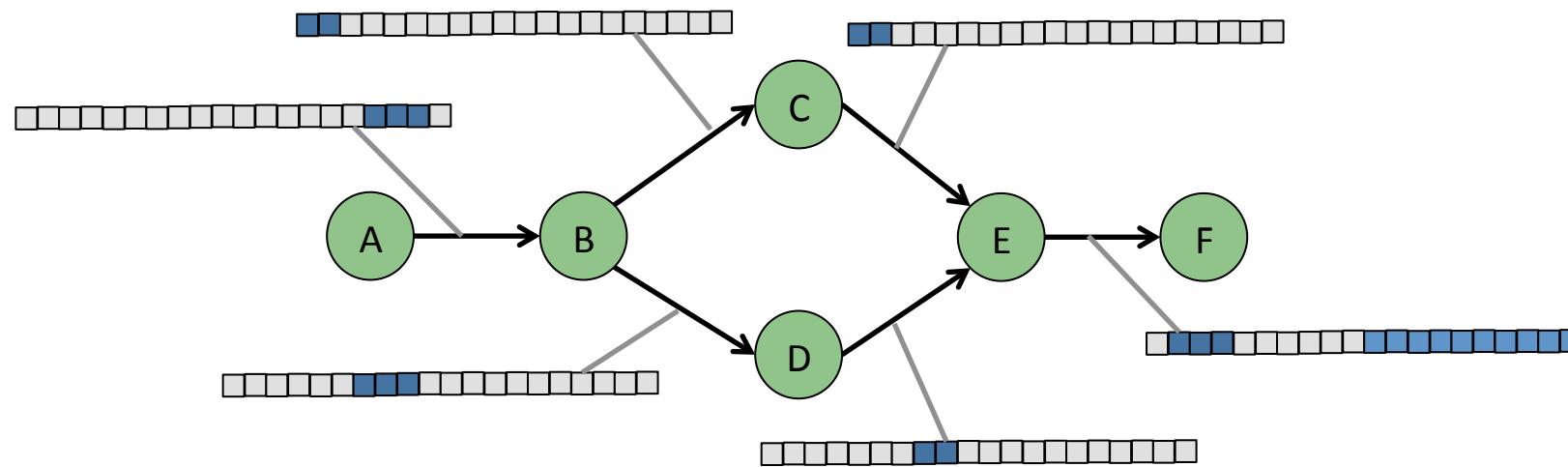
**≠**

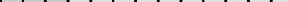


# Temporal Networks

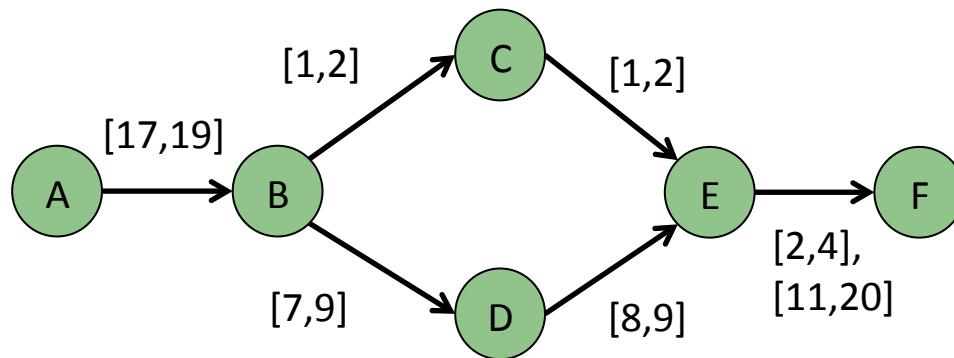


# Temporal Networks

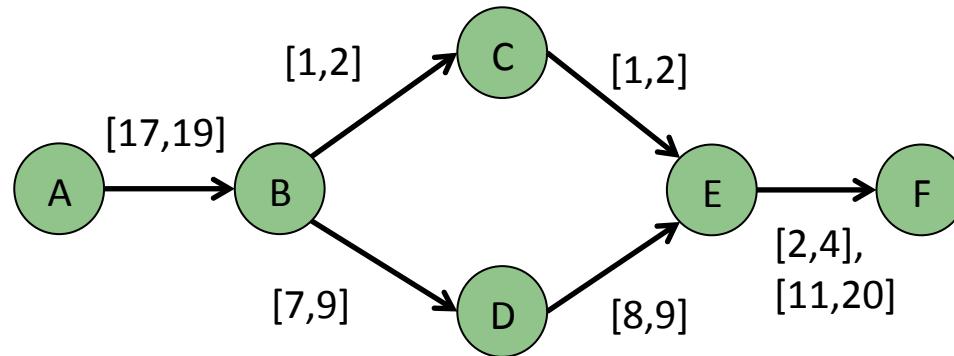


ObServation window W = 

# Temporal Networks

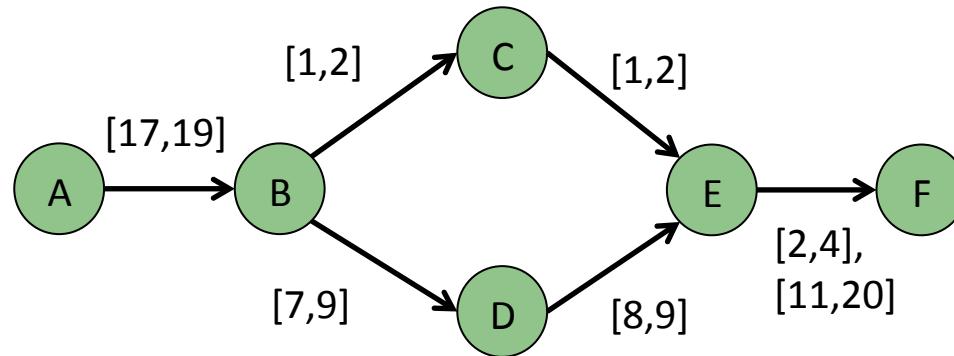


# Temporal Networks



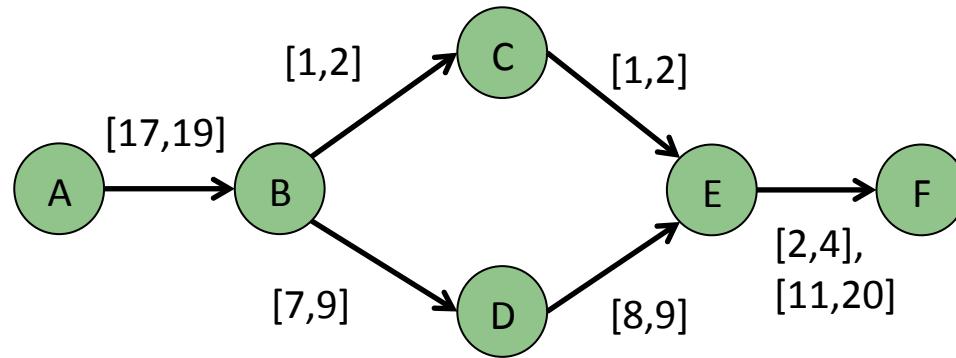
- ✓ PerSon-to-perSon communication
- ✓ DiSease Spreading
- ✓ Social networks
- ✓ ...

# Temporal Networks



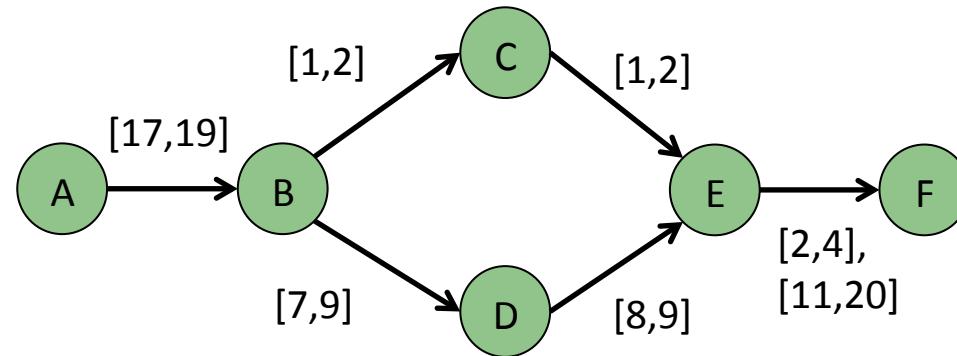
- ✓ PerSon-to-perSon communication
- ✓ DiSease Spreading
- ✓ Social networks
- ✓ ...

# Twitter as a Temporal Network

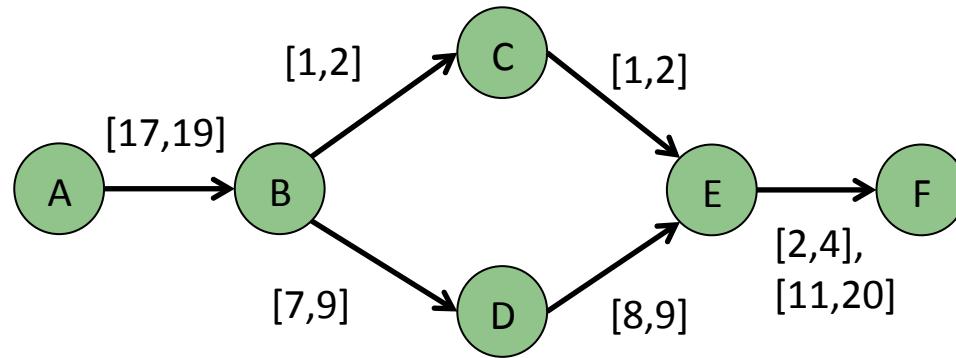


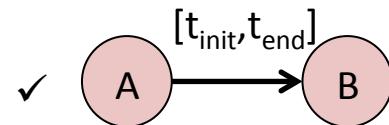
Why are we using temporal networks?

# Twitter as a Temporal Network

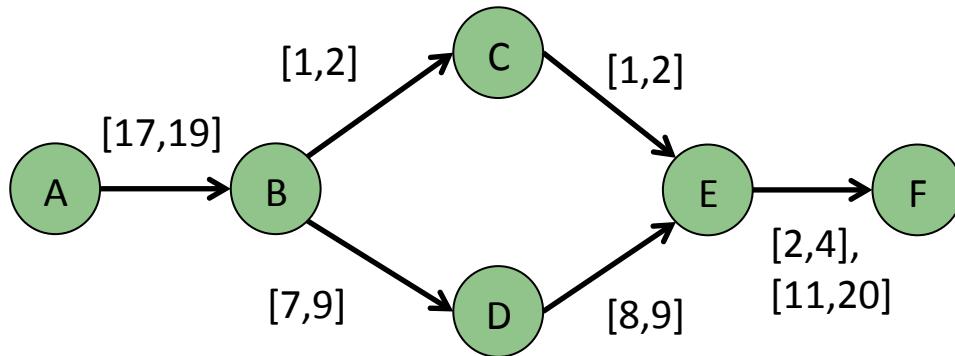


# Twitter as a Temporal Network



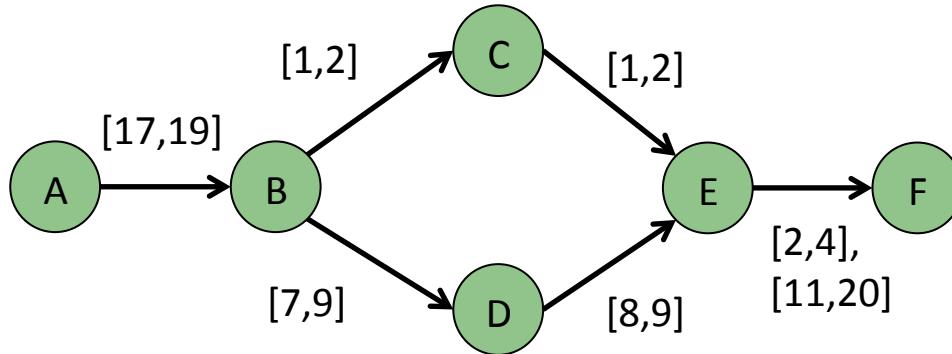
✓  B follows A (information flow) from  $t_{\text{init}}$  to  $t_{\text{end}}$

# Twitter as a Temporal Network



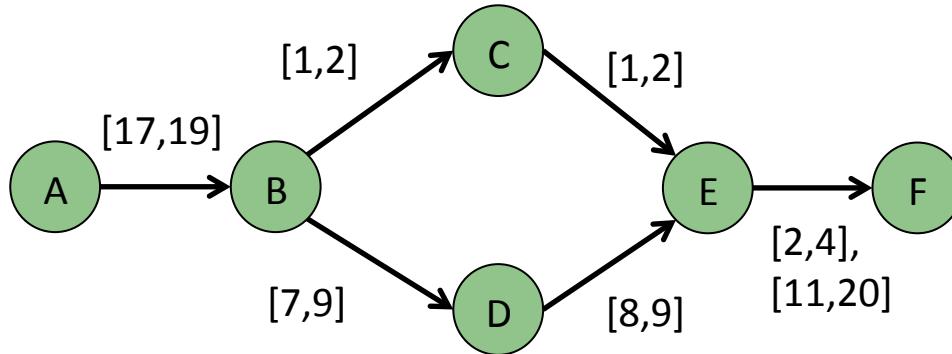
- ✓ B follows A (information flow) from  $t_{init}$  to  $t_{end}$
- ✓  $W = [n, N] = \text{observation window}$

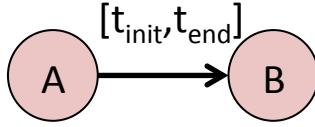
# Twitter as a Temporal Network



- ✓ B follows A (information flow) from  $t_{init}$  to  $t_{end}$
- ✓  $W = [n, N] = \text{observation window}$
- ✓  $R = \text{retention time} = 1 \text{ day}$

# Twitter as a Temporal Network

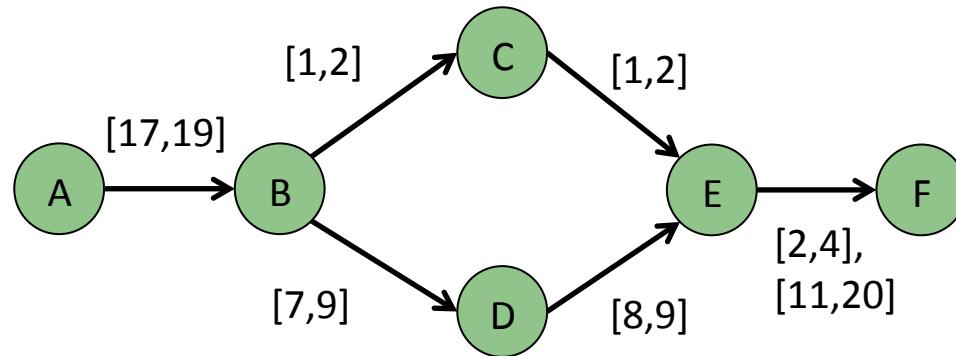


- ✓  B follows A (information flow) from  $t_{init}$  to  $t_{end}$
- ✓  $W = [n, N] = \text{observation window}$
- ✓  $R = \text{retention time} = 1 \text{ day}$
- ✓  $T = \text{edge traversal time} = 0$

# Temporal Metrics

*The concept of geodesic distance  
cannot be limited to the number of  
hops separating two nodes but should  
also take into account the  
temporal ordering of links.*

# Temporal Path



$W = [1,20]$   
 $R = 1 \text{ day}$   
 $T = 0$

Temporal Path	Duration	Is fastest path?
$P(B,E) = <(B,C,1), (C,E,2)>$	1	yes
$P(B,E) = <(B,D,7), (D,E,9)>$	2	no
$P(B,E) = <(B,D,7), (D,E,8)>$	1	yes

# Temporal Metrics

Revisiting centrality metrics...

$$closeness(v) = \sum_{u \in V \setminus \{v\}} \frac{1}{\min(d_{P_{v,u}})}$$

Fastest path duration

- ✓ How close a node is from the other nodes in the graph
- ✓ High closeness = best visibility into what is happening

# Temporal Metrics

Revisiting centrality metrics...

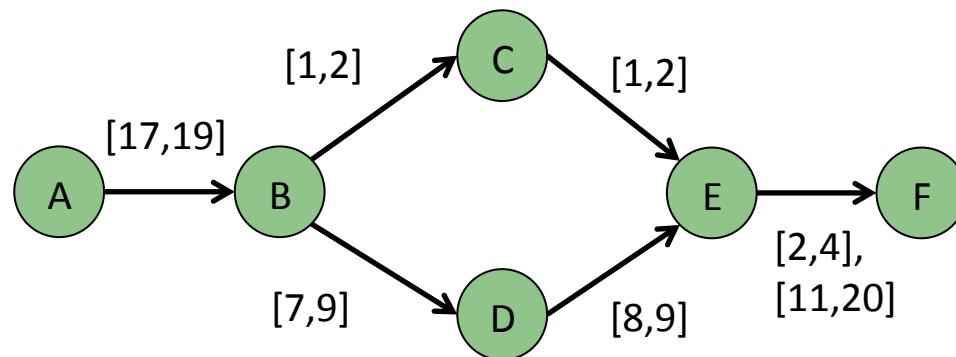
$$\text{betweenness}(v) = \sum_{v \neq j \neq k} \frac{w_v(j, k)}{w(j, k)}$$

Number of fastest paths between j and k that pass through v

Number of fastest paths between j and k

- ✓ High betweenness = great influence over what flows
- ✓ High betweenness = control the flow of information (gatekeeper)

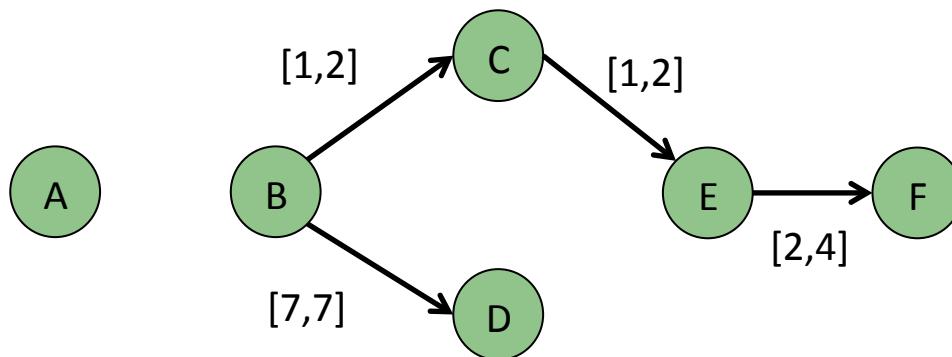
# Betweenness centrality



$W = [1,20]$   
 $R = 1$  day  
 $T = 0$

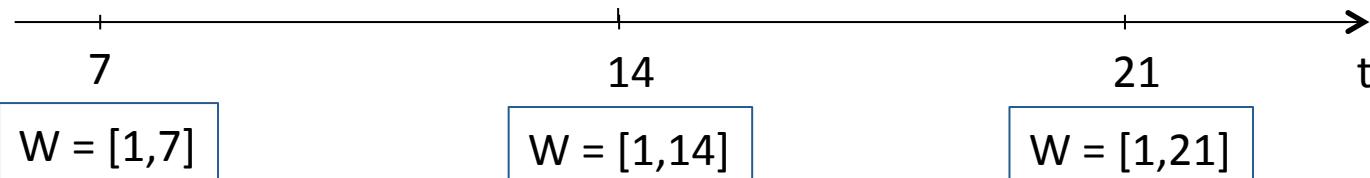
Node	Fastest Path (temporal)	Shortest Path (static)
C	1.33	2
D	0.66	2
E	3	4
B	0	4

# Betweenness centrality

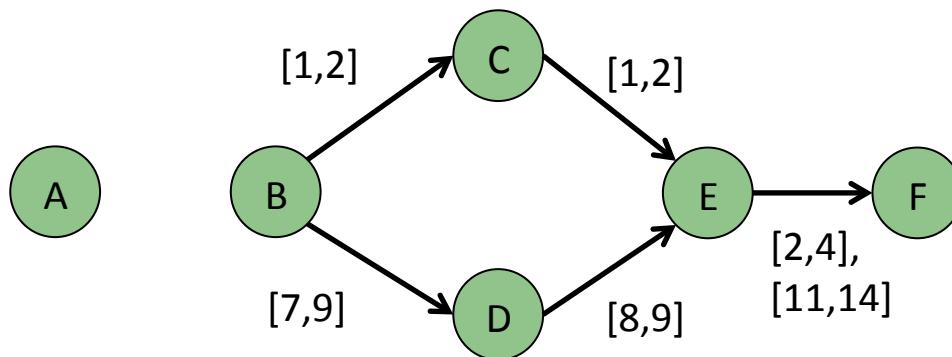


R = 1 day  
T = 0

Ranking	Node
1	E, C
2	A, B, D, F
3	-
4	-



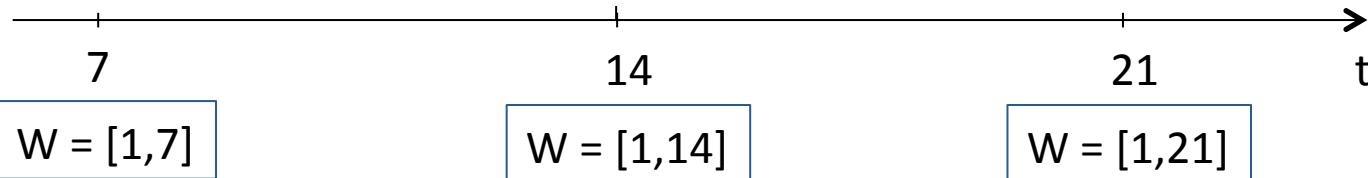
# Betweenness centrality



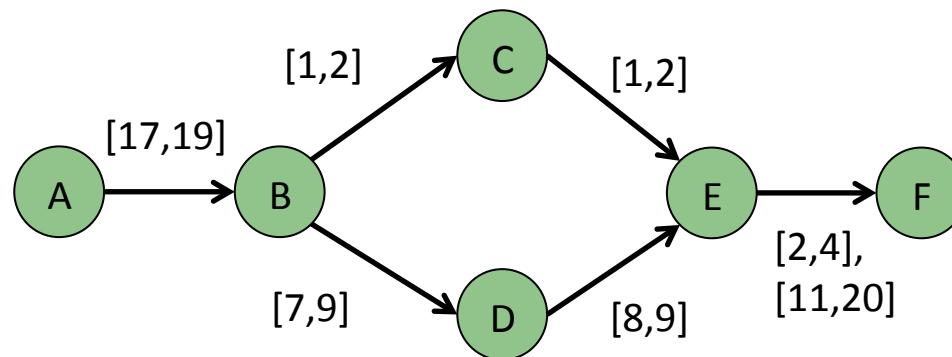
R = 1 day  
T = 0

Ranking	Node
1	E, C
2	A, B, D, F
3	-
4	-

Ranking	Node
1	E
2	C
3	D
4	A, B, F



# Betweenness centrality

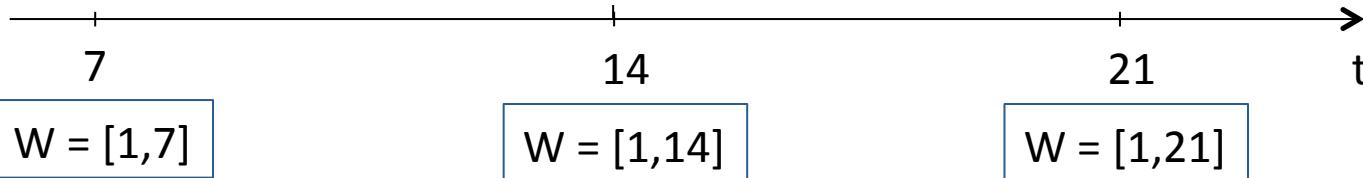


R = 1 day  
T = 0

Ranking	Node
1	E, C
2	A, B, D, F
3	-
4	-

Ranking	Node
1	E
2	C
3	D
4	A, B, F

Ranking	Node
1	E
2	C
3	D
4	A, B, F



# Centrality Change

$$C_i(t) = | \text{pos}_i(t-1) - \text{pos}_i(t) | / \max(\text{pos}_i(t-1), \text{pos}_i(t))$$

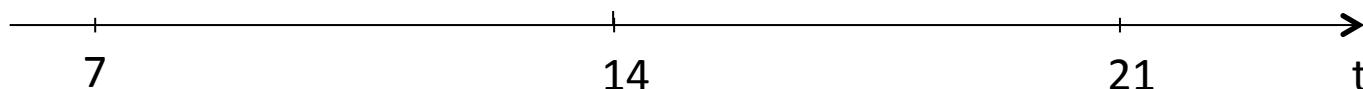


ranking position of  
node  $i$  in time  $t$

Ranking	Node
1	E, C
2	A, B, D, F
3	-
4	-

Ranking	Node
1	E
2	C
3	D
4	A, B, F

Ranking	Node
1	E
2	C
3	D
4	A, B, F



# Outline

Overview

SNA  
Essentials

Temporal  
Social  
Networks

Processing  
Evolving  
Networks

Summary  
of  
Definitions

Case Study  
1: Twitter

Case Study  
2: Calls

Case Study  
3: Last.fm

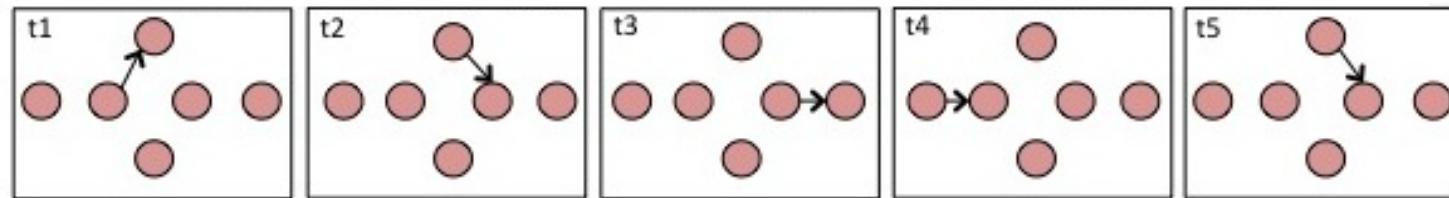
Being a  
Data  
Scientist

Final  
Remarks

# Processing Evolving Networks

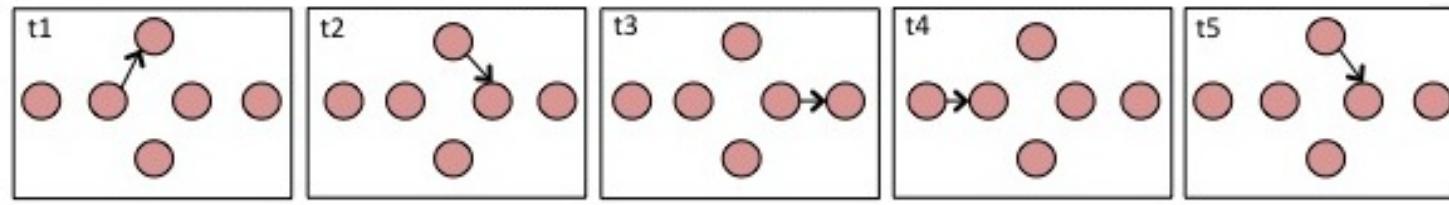
# ProceSSing Evolving Networks

Edges stream (arrival time = time granularity)



# Batch Processing Evolving Networks

Edges stream (arrival time = time granularity)



t1

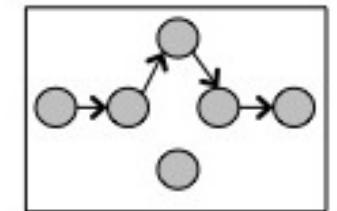
t2

t3

t4

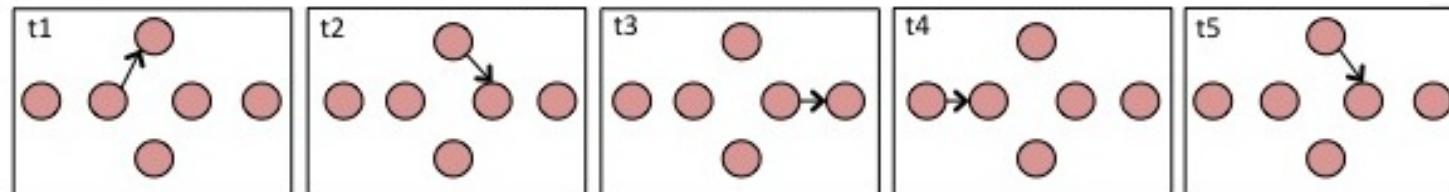
t5

Batch without  
temporal info



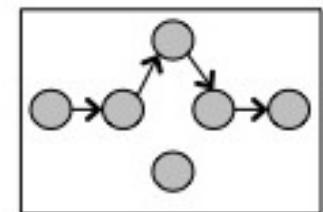
# Batch Processing Evolving Networks

Edges stream (arrival time = time granularity)



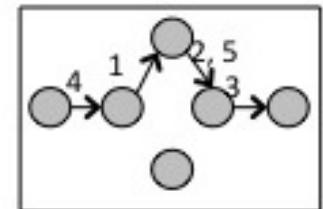
---

Batch without  
temporal info



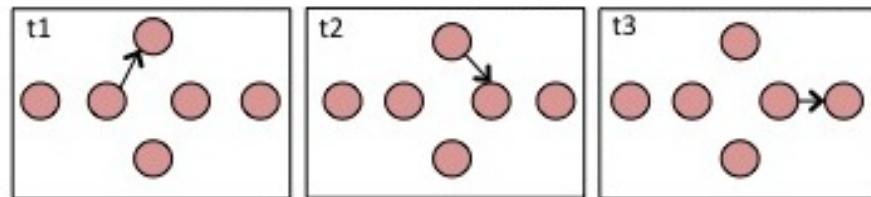
---

Batch with  
temporal info



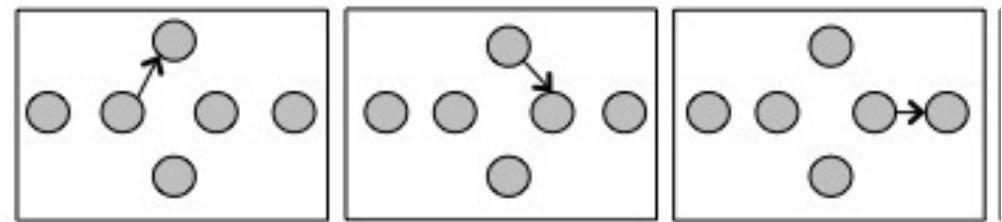
# Batch Processing Evolving Networks

Edges stream (arrival time = time granularity)



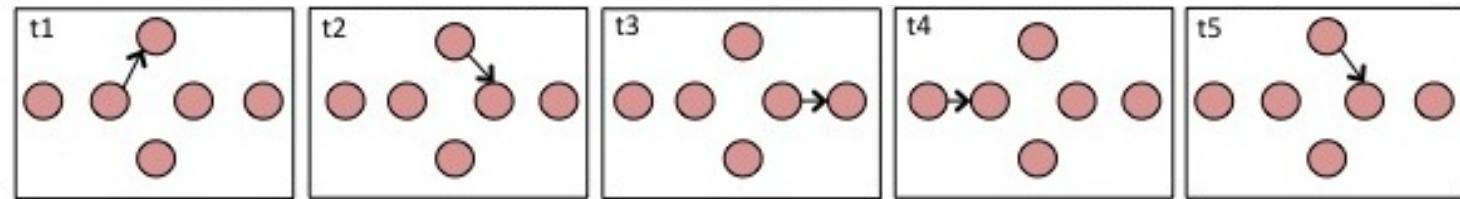
---

batch  
temporal  
snapshots  
(not online)



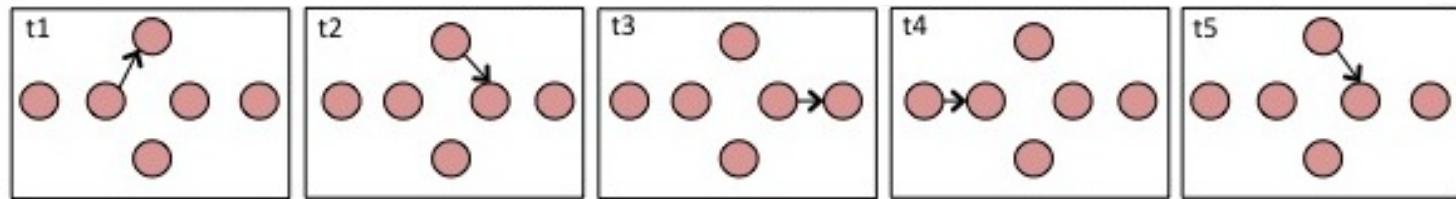
# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)



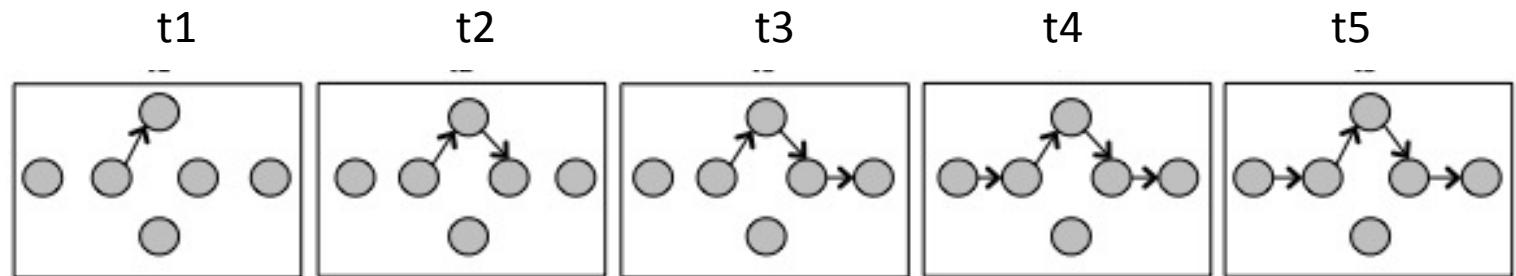
# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)



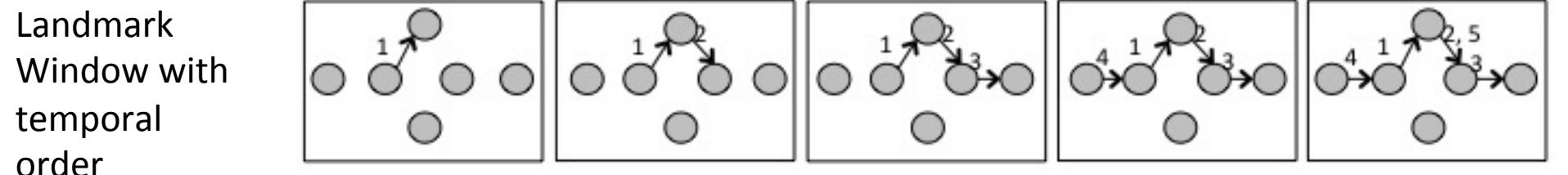
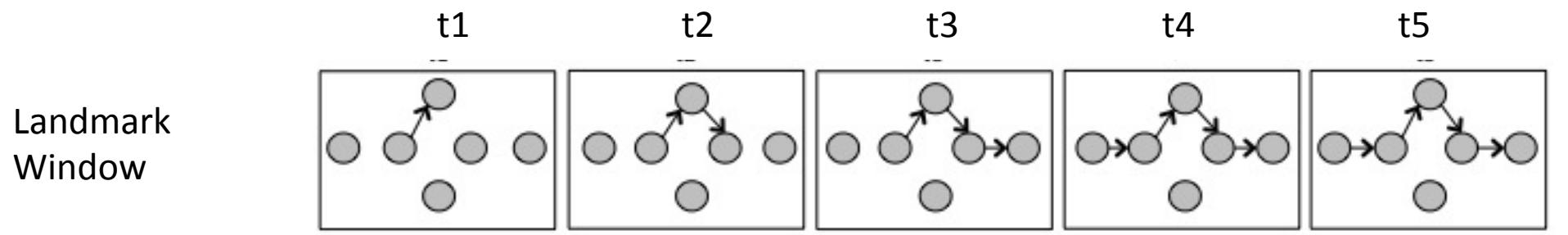
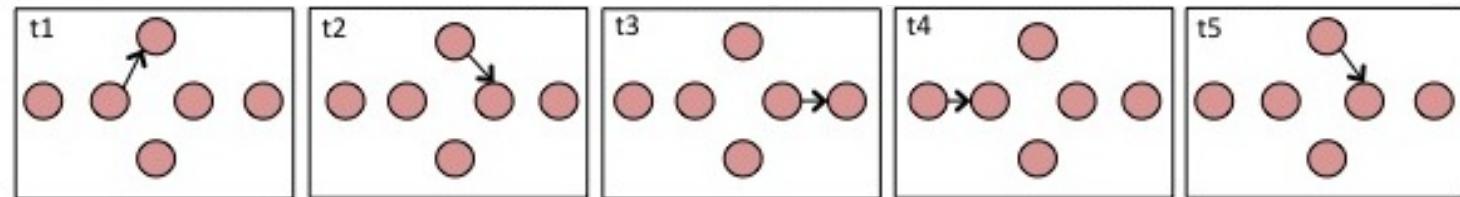
---

Landmark  
Window



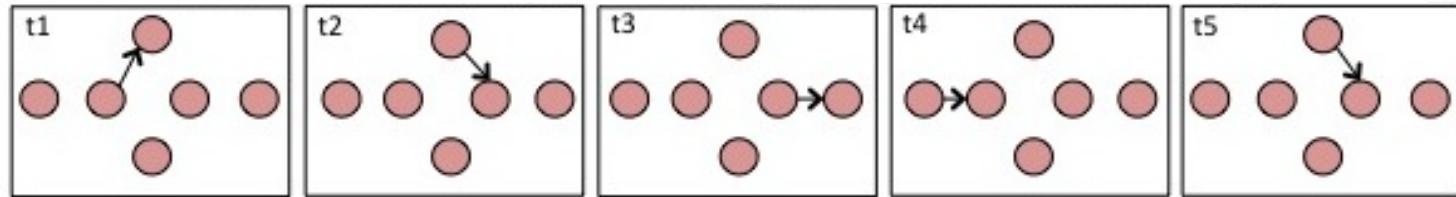
# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)



# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)



---

t1

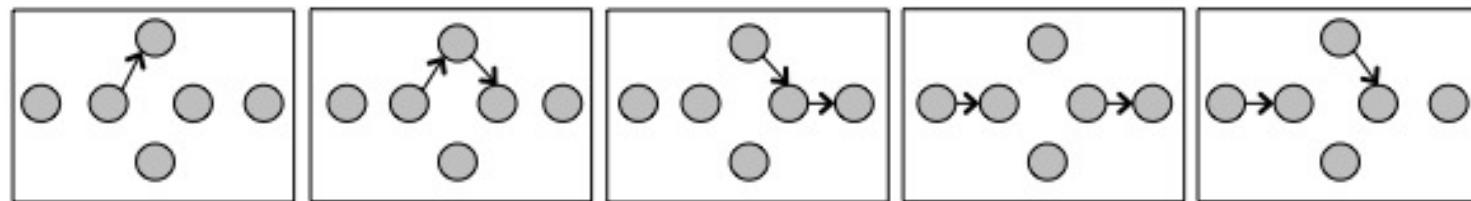
t2

t3

t4

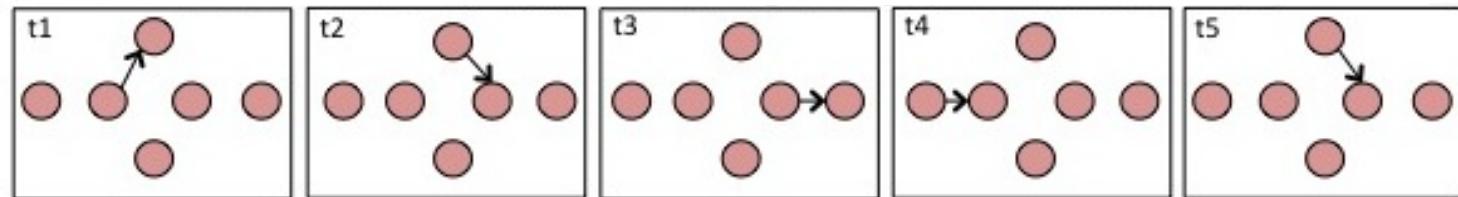
t5

Sliding  
Window  
 $|W| = 2$



# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)

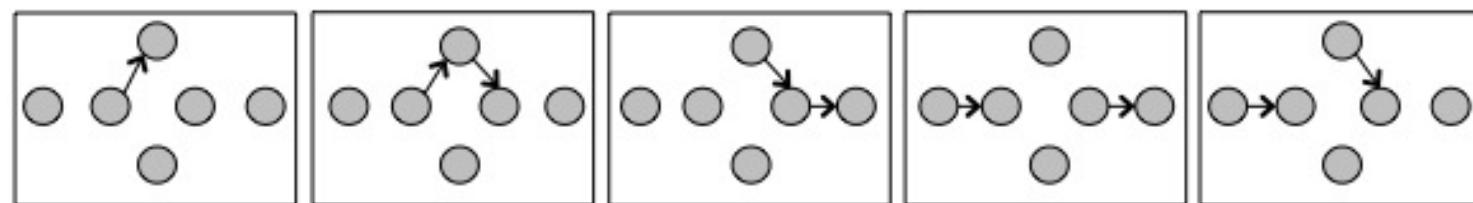


---

t1                    t2                    t3                    t4                    t5

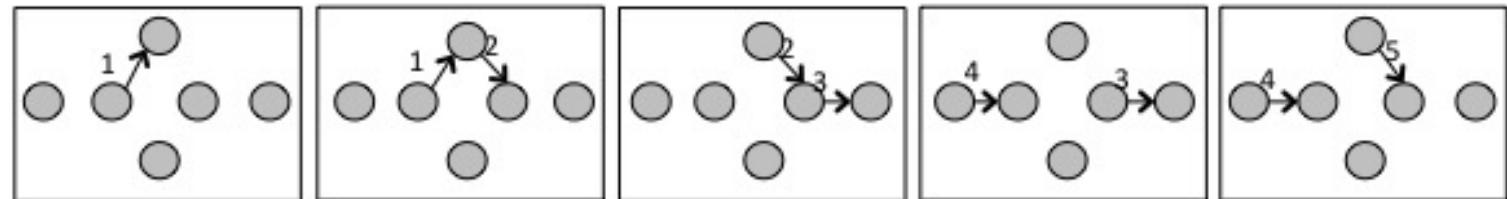
---

Sliding  
Window  
 $|W| = 2$



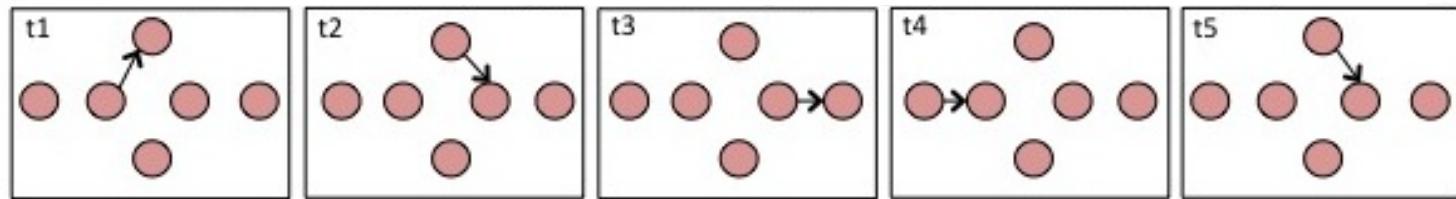
---

Sliding window  
 $|W| = 2$  with  
temporal order



# Online Processing Evolving Networks

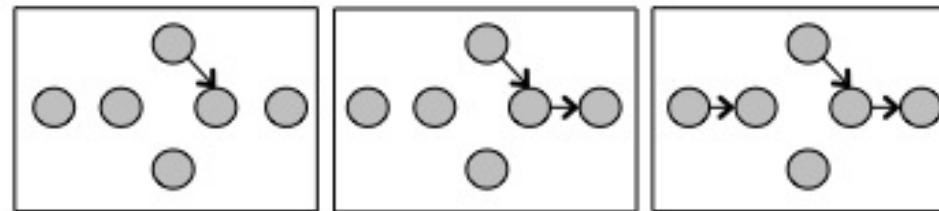
Edges stream (arrival time = time granularity)



---

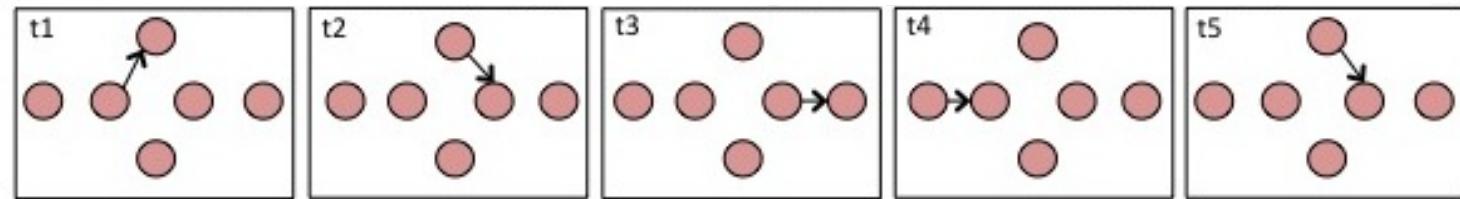
t1                    t2                    t3                    t4                    t5

Fixed  
Observation  
Window  
[t2,t4]



# Online Processing Evolving Networks

Edges stream (arrival time = time granularity)



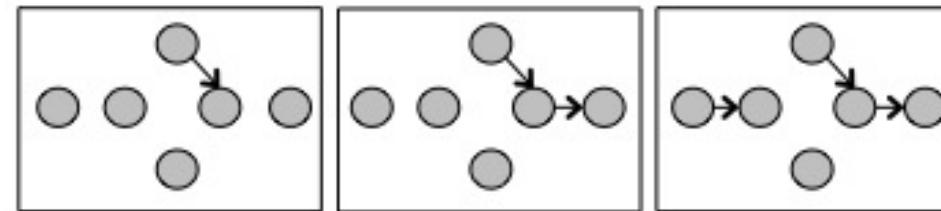
---

t1                    t2                    t3                    t4                    t5

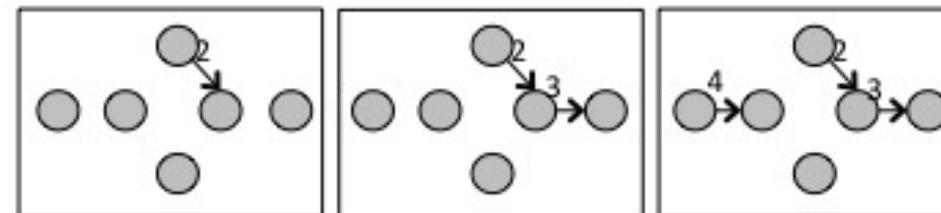
---

Fixed  
Observation  
Window  
[t2,t4]

---

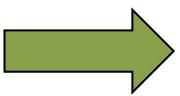


Fixed Observation  
Window [t2,t4]  
with temporal  
order



# Processing Evolving Networks

Batch  Slowly evolving networks

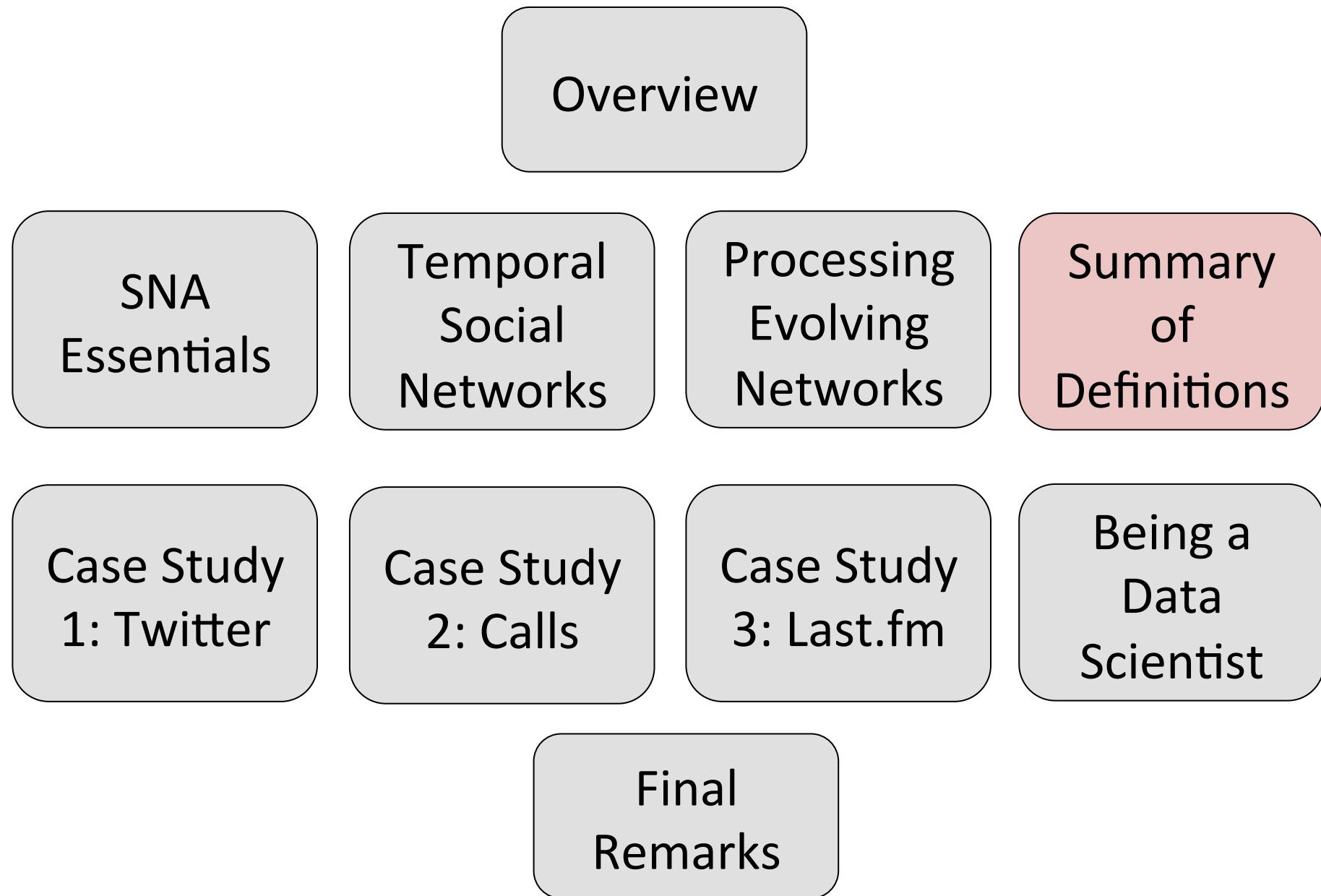
Online  Streaming networks

# HANDS ON!!

[www.lsi.facom.ufu.br/~fabiola/sbbd2017](http://www.lsi.facom.ufu.br/~fabiola/sbbd2017)



# Outline



# Leveraging Definitions

# Definitions

## Network

graph with data in nodes and/or edges (semantics)

## Social Network

Social interaction represented as network.

Property of nonrandomness – relationships tend to cluster (Ullman 2011)

# Definitions

{Evolving, Evolutionary, Time-evolving,  
Dynamic, Time-varying} Network

networks that are changing, with nodes/edges  
appearing and disappearing, associating and  
disassociating as time flies

# Definitions

## Temporal Network

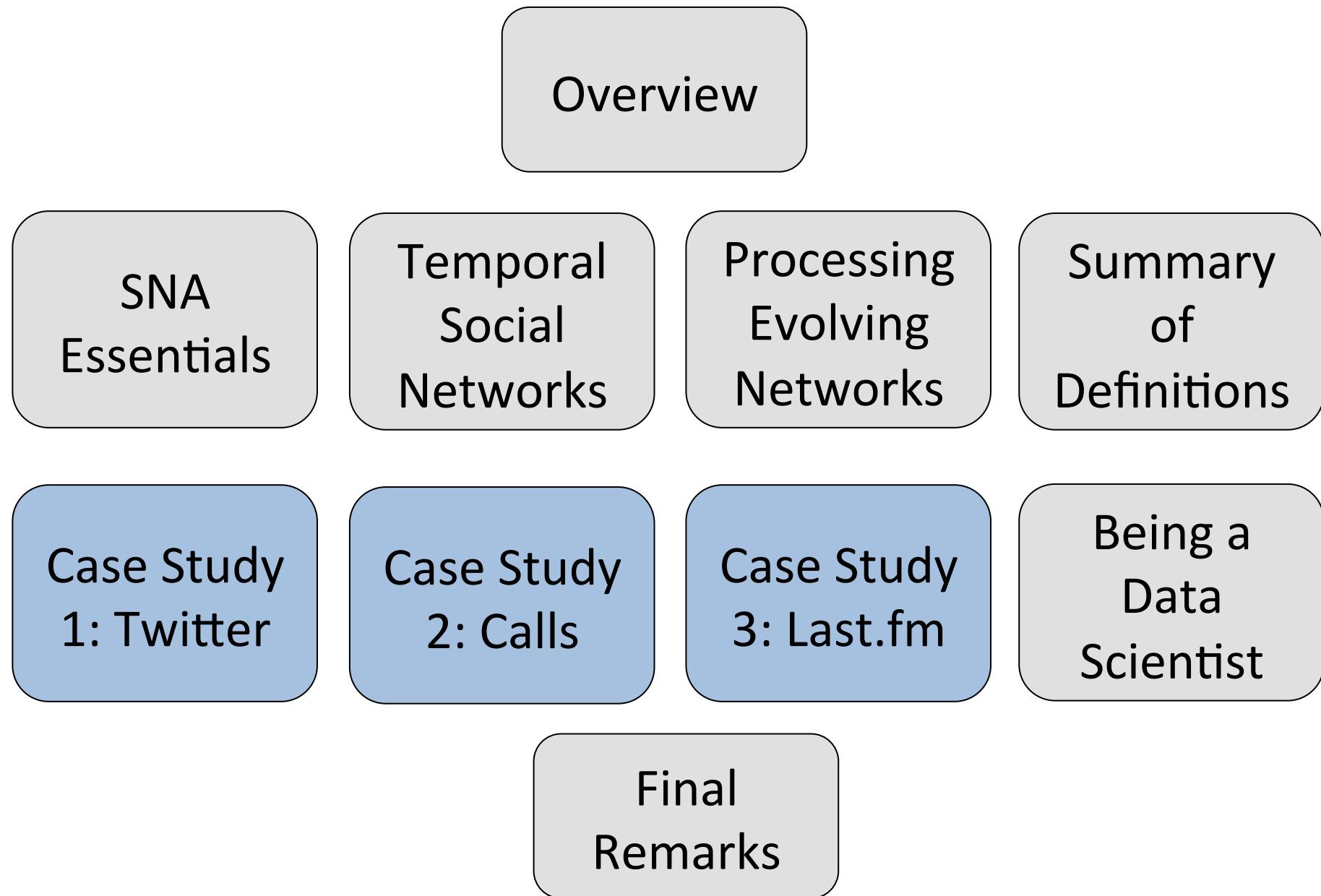
the order that edges appear and disappear is taken into account during the analysis – the temporal order

# Definitions

## Streaming Network

refers to the way that the network is observed and processed, especially in scenarios where there is not the notion of begin and end of the network

# Outline



# Case Study 1

Twitter Social Network

(Pereira et al 2016)

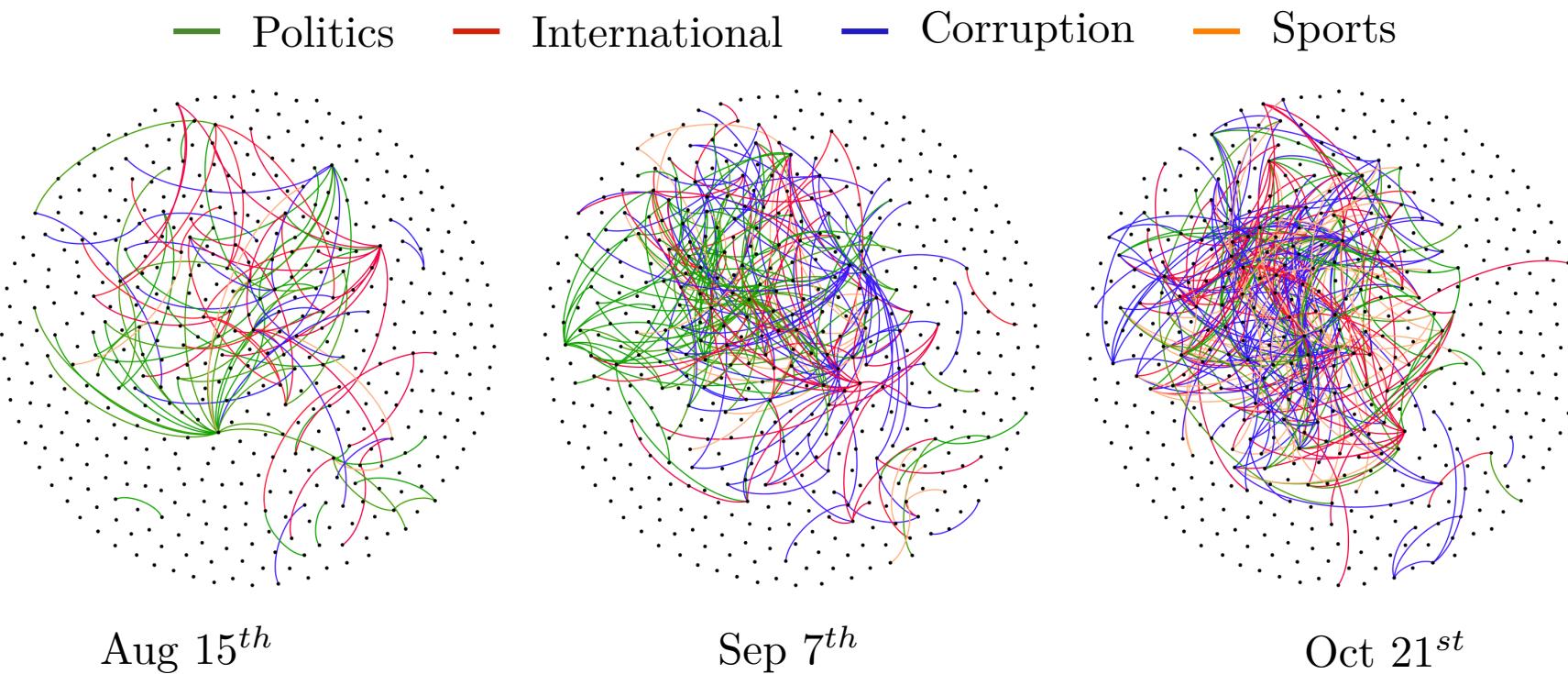
# Network Statistics

Network content	
Domain	Brazilian news in Twitter
Crawling policy	tweets with @folha mention
Time span of tweets posting times	Aug 08 2016 - Nov 09 2016
# total collected tweets	1,771,435
# tweets retweeted	150,822
Network topology	
# nodes	292,310
# temporal edges (retweets)	1,392,841
Avg static path length	12.31
Avg temporal path length	5 days
Avg static degree	4.76

# Centrality Metrics

<b>Network metrics</b>	<b>static</b>	<b>temporal</b>
Average shortest path	12.31	5 days
Average degree	4.76	-
Average closeness	1.01	2.44
Average betweenness	0.0056	0.0233

# Giving a meaning to the SNA



**Fig. 4** Snapshots of samples of the evolving interaction network. Nodes are Twitter users. One tie from user  $u_1$  to  $u_2$  means that  $u_2$  retweeted at  $t$  some text originally posted by  $u_1$ . The colors represent topics that users are talking about at  $t$ . The samples were built by filtering nodes with degree between 50-22000 and edges representing the 4 most popular topics. Each snapshot corresponds to 1 day time-interval. This figure highlights the *edges* evolving aspect. Nodes are not evolving for better visualization.

# The findings

User type based on retweet behavior:  
producer, consumer and  
producer&consumer

Predominant topics vary over time

Users favorite topics just analyzing  
the network

# Case Study 2

Telecom Calls Network

(Tabassum and Gama 2016)

# Network Statistics

## Network description

---

Domain	Telecom Calls
Time interval	31 days
Processing Granularity	1 day
Processing Type	streaming (online)

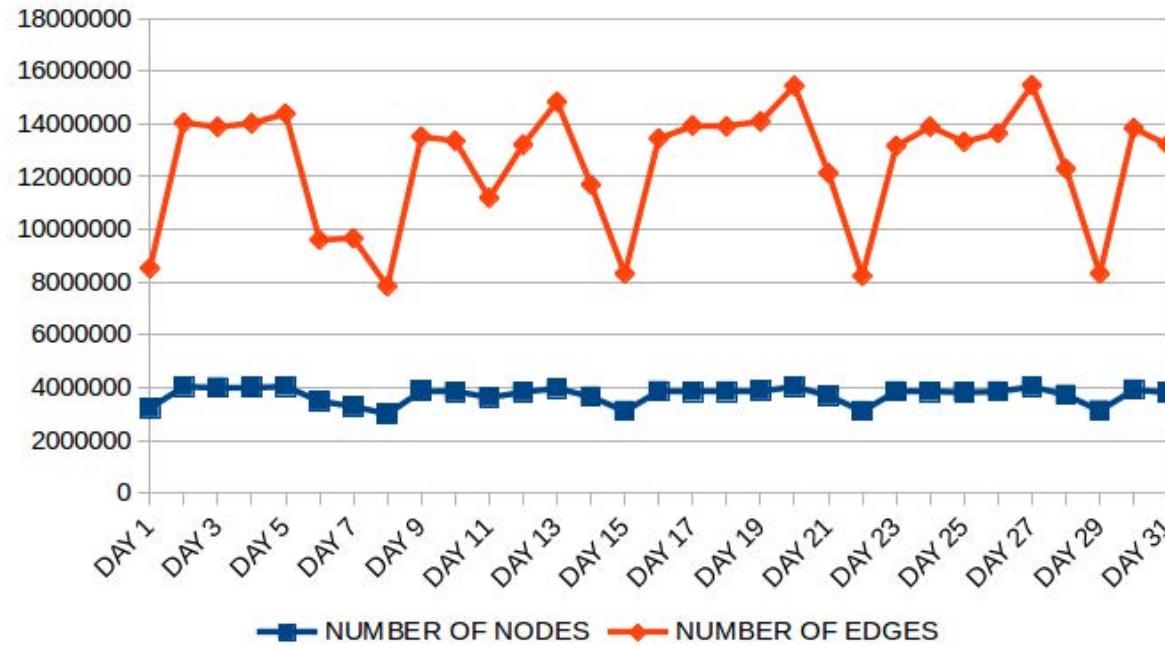
## Network topology

---

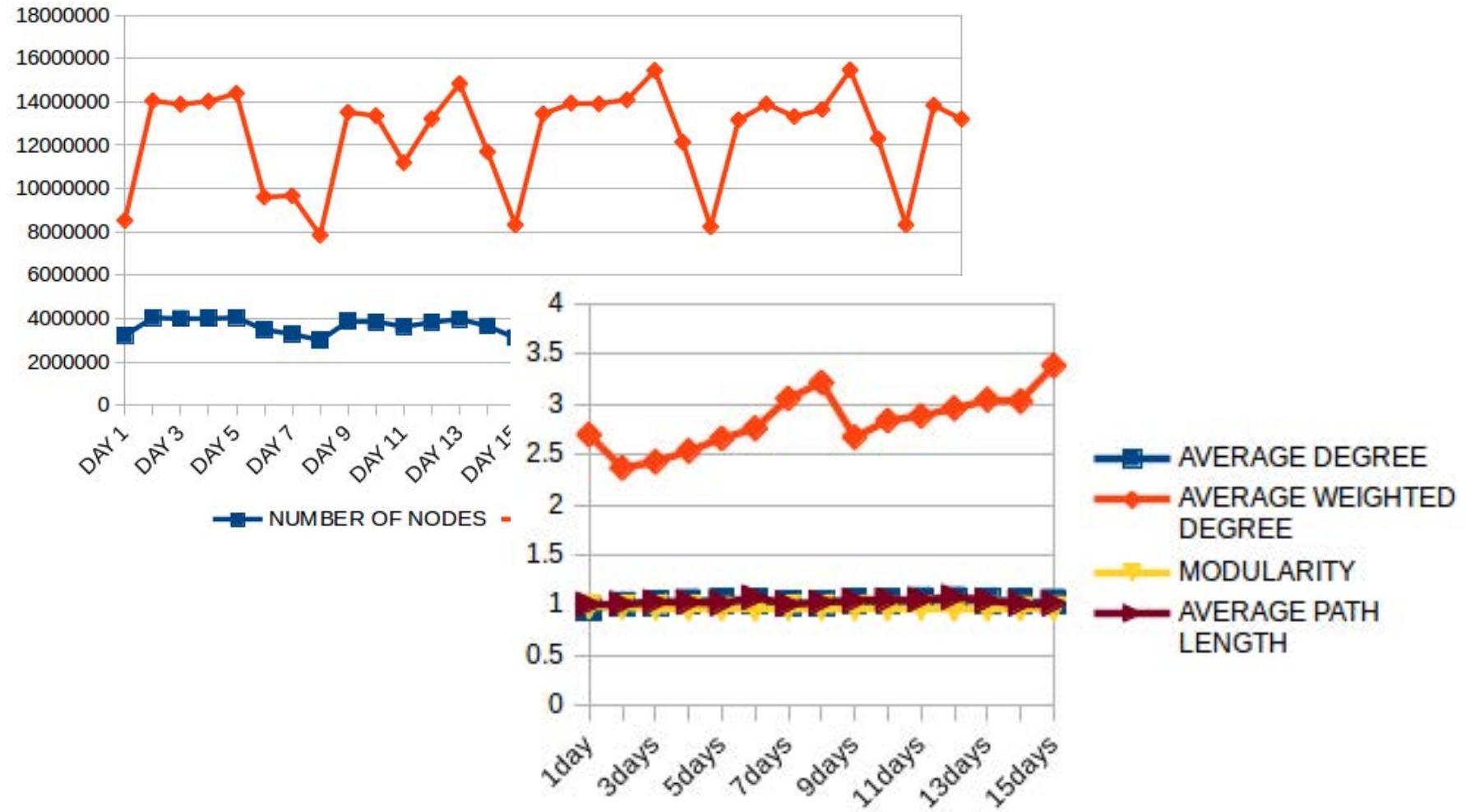
# calls (edges)	386,492,749
# phone numbers (nodes)	11,916,442

---

# Network Description



# Network Description



# Biased Random Sampling Algorithm

**input** : Unbounded *stream*

**output**: Realtime *sample* of size *k*

*Filling the reservoir with first k items/objects;*

**for** *i* = 1 **to** *k* **do**

| *sample*[*i*]  $\leftarrow$  *stream*[*i*];

**end**

*Inserting all new items into the stream;*

**while** *stream*!= *EOF* **do**

| *i* = *i* + 1;

| *pos*  $\leftarrow$  *Random*(1, *k*);

| *sample*[*pos*]  $\leftarrow$  *stream*[*i*];

**end**

# The Findings

Online processing the evolving network  
using sampling strategy

There is a week-pattern of calls

Constant metrics: telecom company clients  
are stable

# Case Study 3

Last.fm Music Network

(Palovics et al 2013)

# Network Statistics

## Network description

---

Domain	Friendship in Last.fm
Time interval	1/1/2002 – 12/31/2011
Processing Granularity	1 day

## Network topology

---

# users (nodes)	71,000
# relationships (edges)	285,241

# Network Description

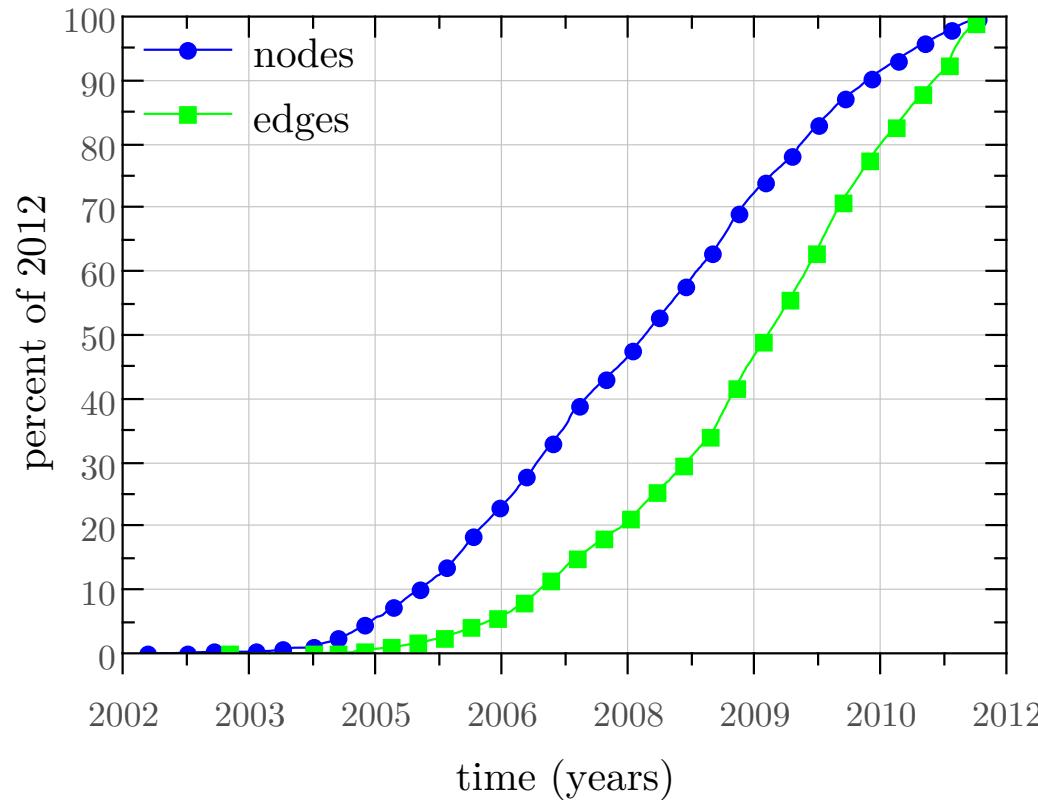


Fig. 1. The number of the users and friendship edges in time as the fraction of the values at the time of the data set creation (2012).

# Network Description

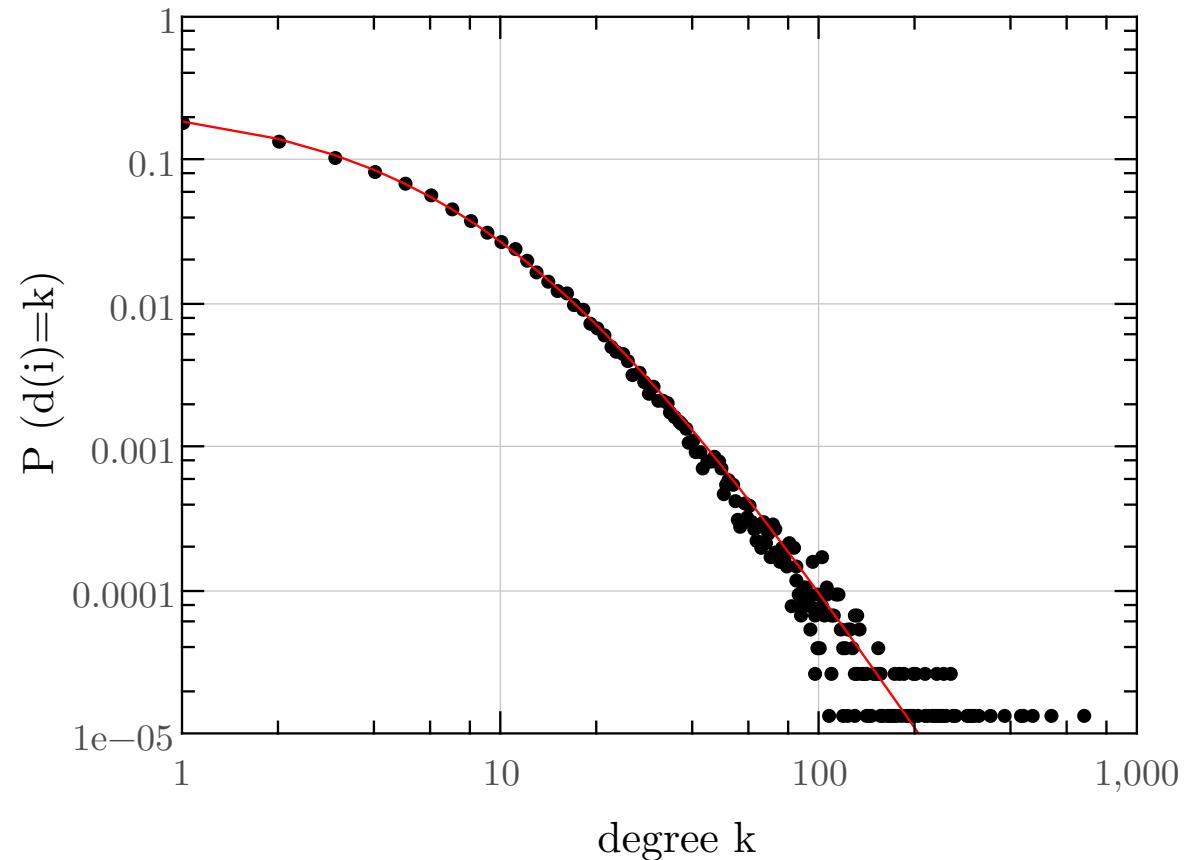
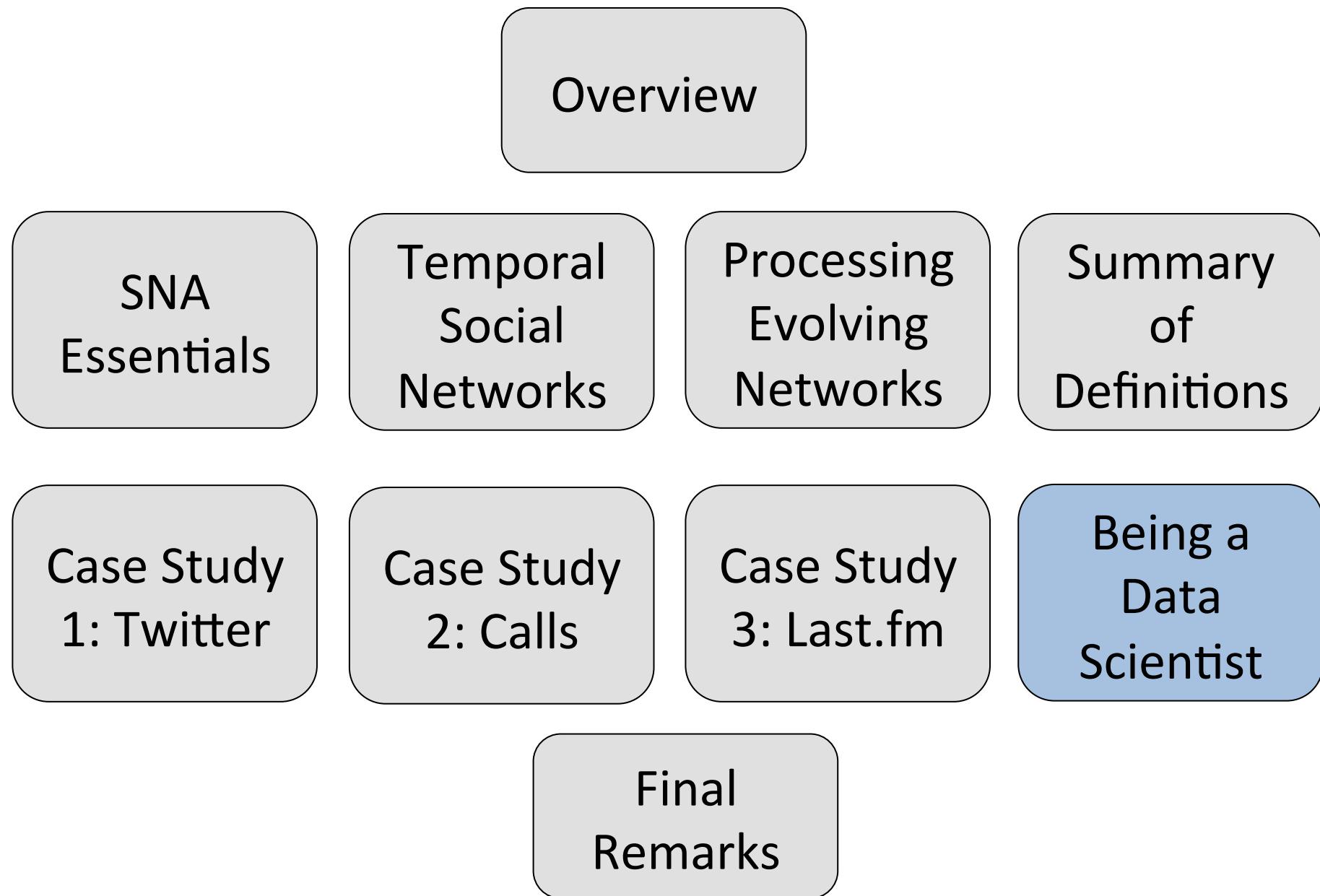


Fig. 3. Degree distribution in the friendship network.

# The Findings

Influence level that friendship can exert  
over music tastes

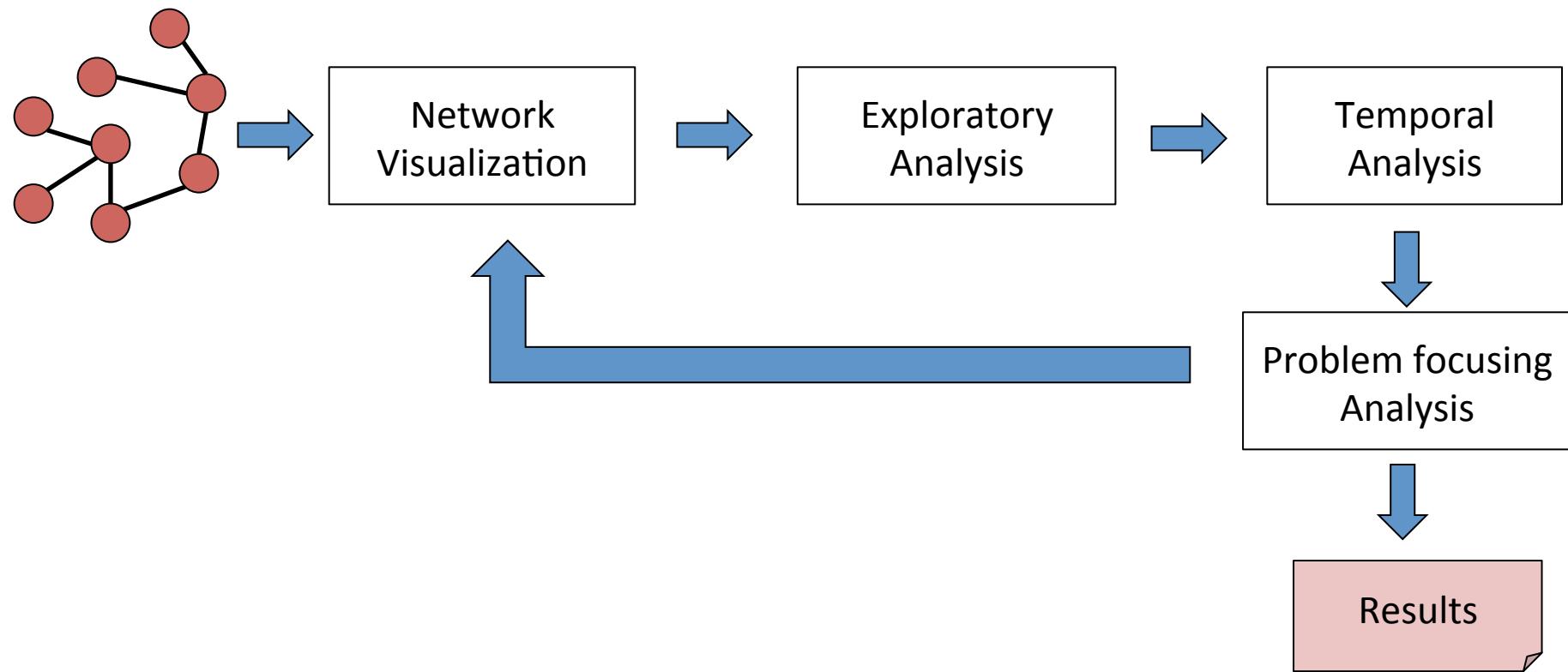
# Outline



Now...

you are a data scientist  
Specialized on temporal  
Social networks!

# Remember...



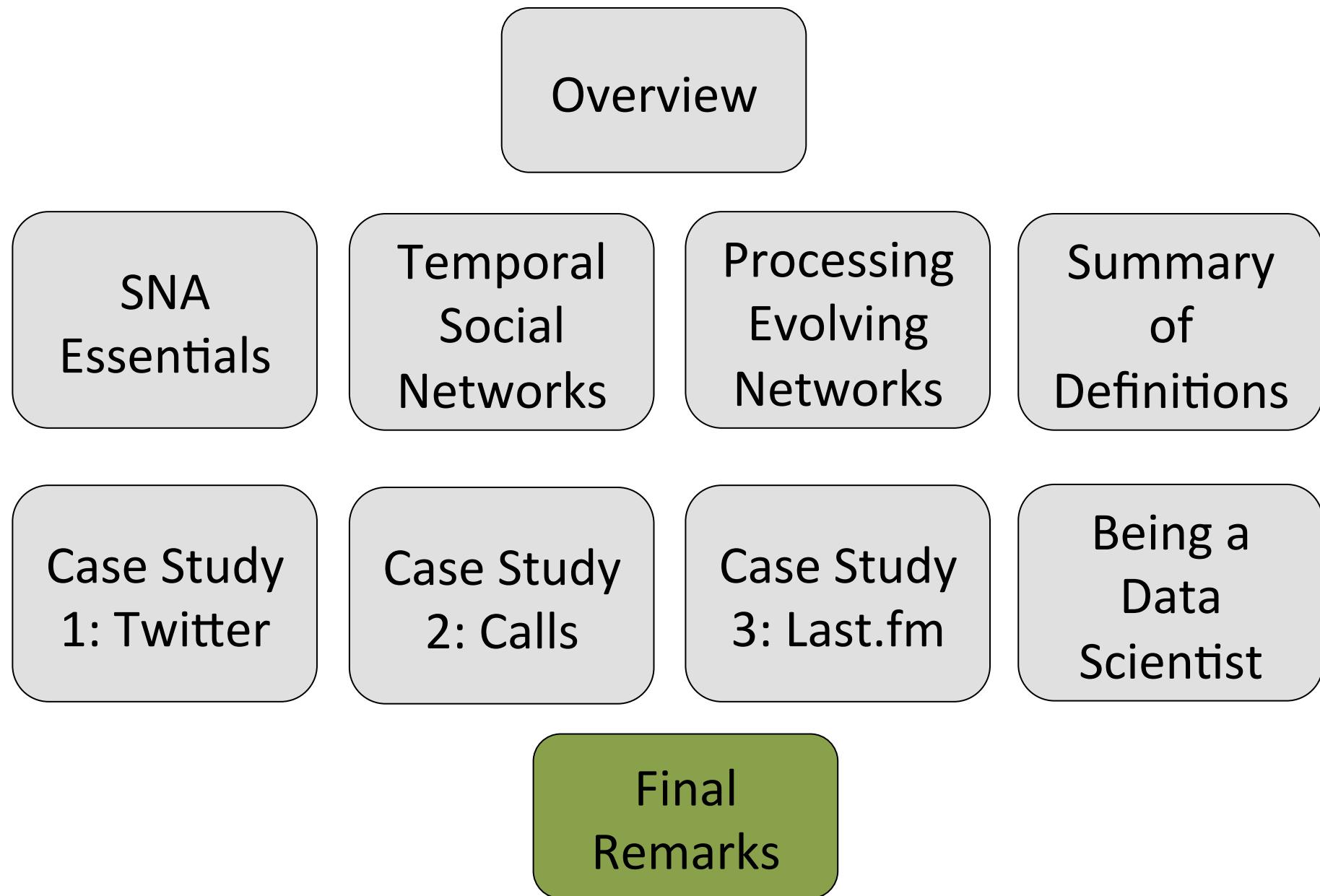
## Steps for Knowledge Discovery from Temporal Social Networks

# HANDS ON!!

[www.lsi.facom.ufu.br/~fabiola/sbbd2017](http://www.lsi.facom.ufu.br/~fabiola/sbbd2017)



# Outline



# Final Remarks

# SNA Tools

- ✓ Gephi (visualization, basic network metrics, temporal networks)
- ✓ iGraph (programming assignments)
  - ✓ <http://igraph.org/r/>
- ✓ NetLogo (modeling network dynamics)

# SNA Tools

- ✓ UCINet
  - ✓ sociology-focused functionality
  - ✓ Windows-only
  - ✓ not-free \$
- ✓ ORA (CMU)
  - ✓ business-focused
  - ✓ Windows-only (.exe)
  - ✓ versions LITE and Professional (\$)

# SNA Tools

- ✓ NodeXL
  - ✓ SNA integrated into Excel
  - ✓ Windows-only
  - ✓ free

# SNA Tools

- ✓ NetworkX (Python)
  - ✓ extensive functionality
  - ✓ scales to large networks
  - ✓ open source

# SNA Tools

- ✓ SNA package for R
  - ✓ extensive, statistics-heavy functionality
- ✓ SoNIA – Social Network Image Animator
  - ✓ specialized for longitudinal analysis of networks
- ✓ SNAP System
  - ✓ <http://snap.stanford.edu/snap/index.html>

# SNA Tools

- ✓ DyNetVis [Linhares et al 2017]
  - ✓ dynamic networks visualization
  - ✓ under development
- ✓ GraphStreamAPI (Java)
- ✓ Pajek
  - ✓ <http://mrvar.fdv.uni-lj.si/pajek/>

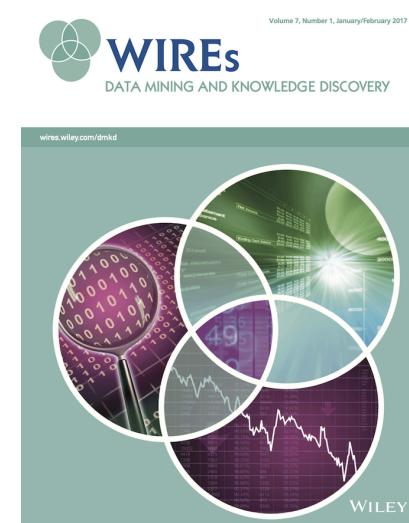
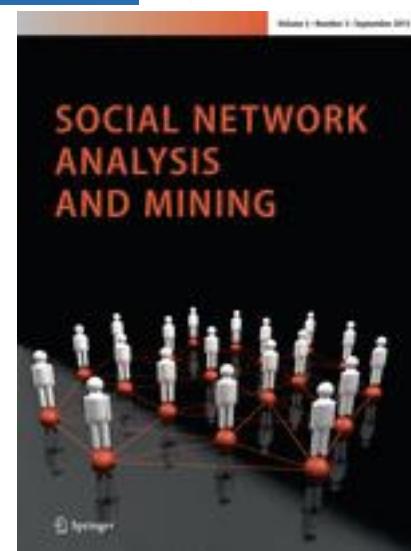
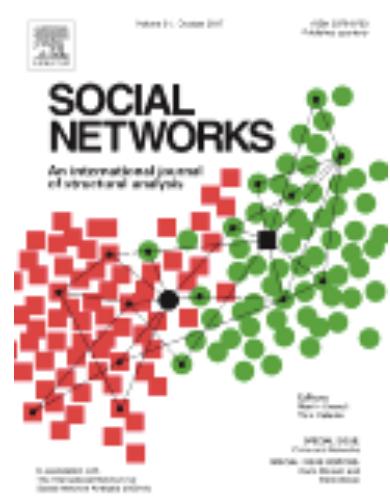
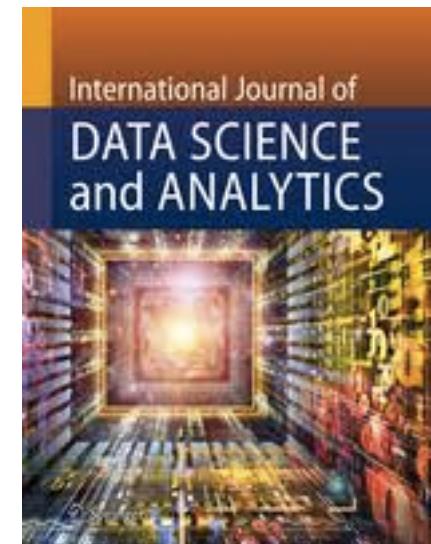
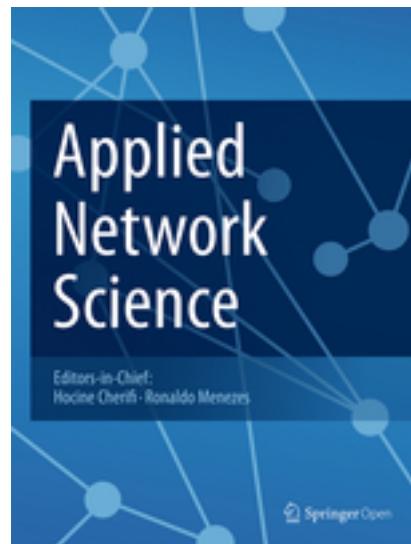
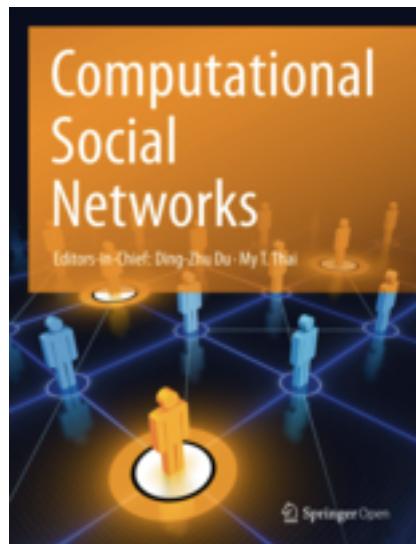
# Top-3 Social Networks Repositories

- ✓ Index of Complex Networks (ICON)
  - ✓ <https://icon.colorado.edu/>
- ✓ SNAP (Stanford Large Network Dataset Collection)
- ✓ KONECT (Koblenz network collection)

# Conferences & Journals

- ICWSM
- BraSNAM
- ASONAM
- ComplexNetworks
- WSDM, HT, ICDM, DSAA, \*KDD, ...

# Conferences & Journals



# Evolving Networks Research Topics

- ✓ Evolutionary network analysis
- ✓ Social networks and social media
- ✓ Graph data analytics
- ✓ Sampling from evolving networks
- ✓ Network data mining
- ✓ Distributed network analysis and mining
- ✓ Statistical techniques for network analysis and mining
- ✓ Temporal and streaming networks
- ✓ Predictive modeling on evolving networks
- ✓ Change detection, anomaly detection
- ✓ Incorporating network content in evolution analysis
- ✓ Community detection in evolving networks

# It's a matter of time!

## Knowledge Discovery from Temporal Social Networks

**Fabíola S. F. Pereira (UFU)**  
João Gama (LIAAD, UPorto)  
Gina M. B. de Oliveira (UFU)

<http://lsi.facom.ufu.br/~fabiola>



# References

## Enrol Mail Dataset (<http://www.enron-mail.com/>)

**[Tang et al. 2010a]** John Tang, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Vincenzo Nicosia. 2010. Analysing information flows and key mediators through temporal centrality metrics. In Proceedings of the 3rd Workshop on Social Network Systems (SNS '10). ACM, New York, NY, USA, , Article 3 , 6 pages.

**[Shetty and Adibi 2005]** Jitesh Shetty and Jafar Adibi. 2005. Discovering important nodes through graph entropy the case of Enron email database. In Proceedings of the 3rd international workshop on Link discovery (LinkKDD '05). ACM, New York, NY, USA, 74-81.

## Facebook Dataset

**[Viswanath et al. 2009a]** Viswanath, B.; Mislove, A.; Cha, M.; and Gummadi, K. P. 2009a. On the evolution of user interaction in facebook. In Proceedings of the 2nd ACM workshop on Online social networks, 37–42. ACM

## Contact Network

**[Cattuto et al. 2013]** Cattuto, C.; Quaggiotto, M.; Panisson, A.; and Averbuch, A. 2013. Time-varying social networks in a graph database: a neo4j use case. In First Interna- tional Workshop on Graph Data Management Experiences and Systems, 11. ACM.

# References

## Epinions Trust Network

[Massa & Avesani 2007] Trust-aware recommender systems. Proceedings of the 2007 ACM Conference on Recommender Systems, 17-24, Minneapolis, USA.

Frank Takes, Social Network Analysis, Challenges in Computer Science – Slides, April 1, 2014 LIACS, Leiden University

Daniel Barbosa, Nuno Filipe, and Joao Gama. A Social Network Analysis of The Portuguese Connection in Panama Papers. FEP Working Papers. ISSN 0870-8541. 2017

Lada Adamic. SNA L1B: software tools.

Linhares, C.D.G., Travençolo, B.A.N., Paiva, J.G.S., Rocha, L.E.C.: Dynetvis: A system for visualization of dynamic networks. In: Proceedings of the Symposium on Applied Computing. pp. 187–194. SAC ’17, ACM, New York, NY, USA (2017)

Pálovics, R., Benczúr, A.A.: Temporal influence over the last.fm social network. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 486–493. ASONAM ’13, ACM, New York, NY, USA (2013),

Pereira, F.S.F., de Amo, S., Gama, J.: Detecting events in evolving social networks through node centrality analysis. Workshop on Large-scale Learning from Data Streams in Evolving Environments co-located with ECML/PKDD (2016)

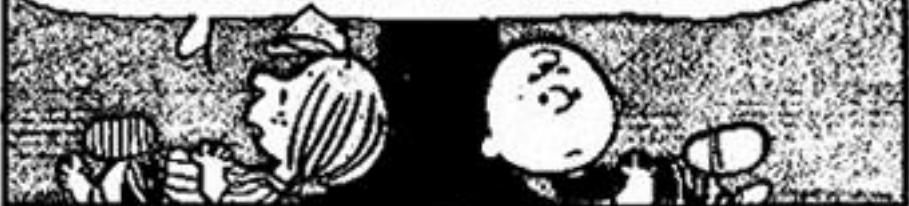
# References

Tabassum,S.,Gama,J.:Sampling massive streaming call graphs. In: Proceedings of the 2016 ACM Symposium on Applied Computing. pp. 923–928. SAC ’16, ACM, New York, NY, USA (2016)

Ullman et al. Mining of Massive Datasets 2011

Bruno Goncalves. Social Network Anlaysis. Lectures Slides. 2014

WELL, IF A LIKES B, BUT B LIKES C  
WHO LIKES D AND E WHO BOTH LIKE A  
WHO DOESN'T EVEN KNOW THAT D EXISTS,  
SHOULD F TRY TO HAVE G TALK TO B SO  
E WILL KNOW THAT C LIKES D AND E, AND  
THAT C WILL POUND H IF SHE COMES  
AROUND AGAIN BUTTING IN ?



MAY I THINK ABOUT THAT  
FOR A MINUTE ?

