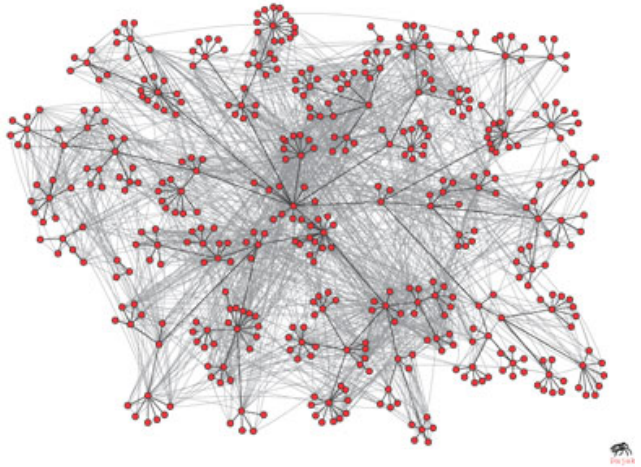
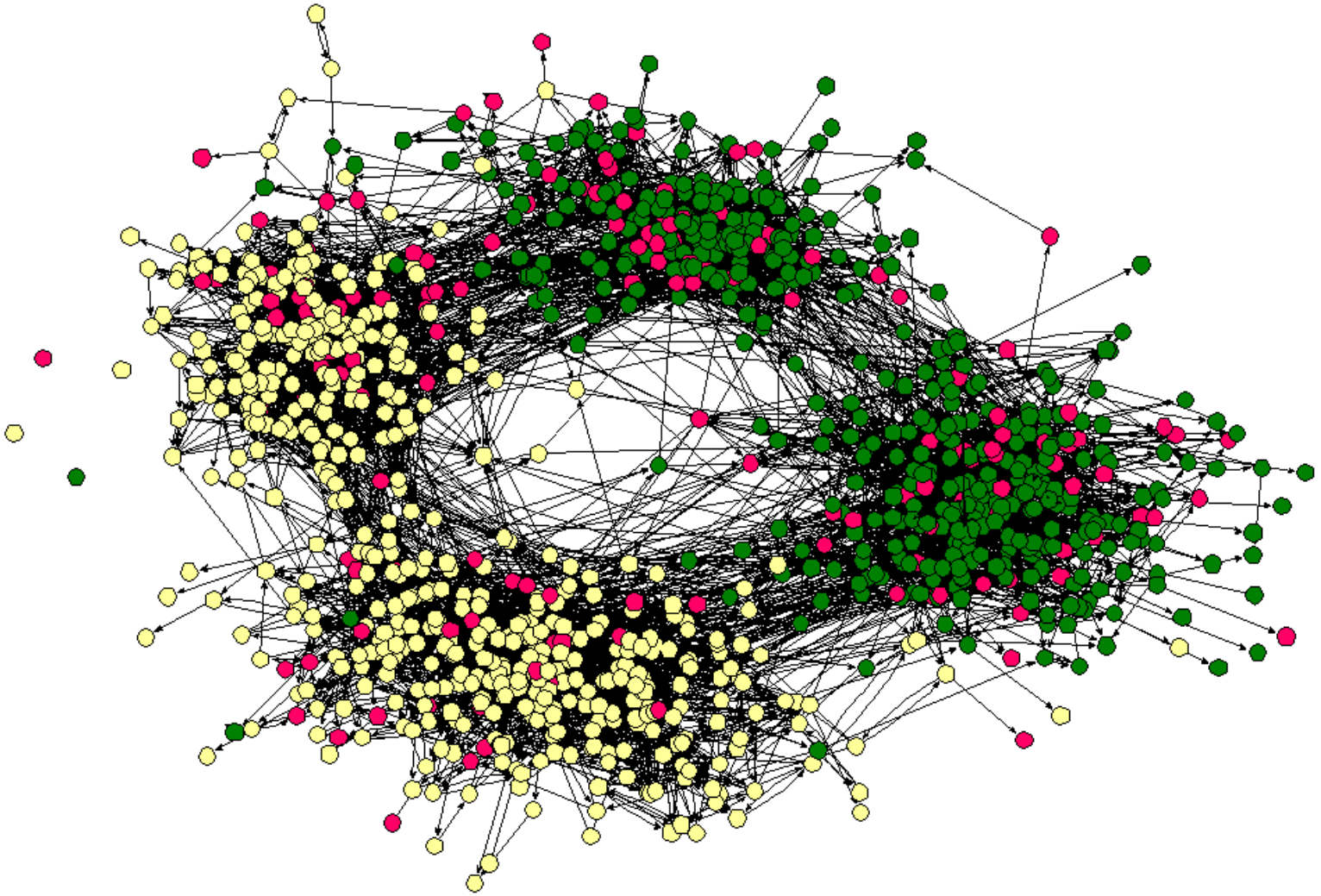
Temporal Graphs for Dynamic Network Analysis

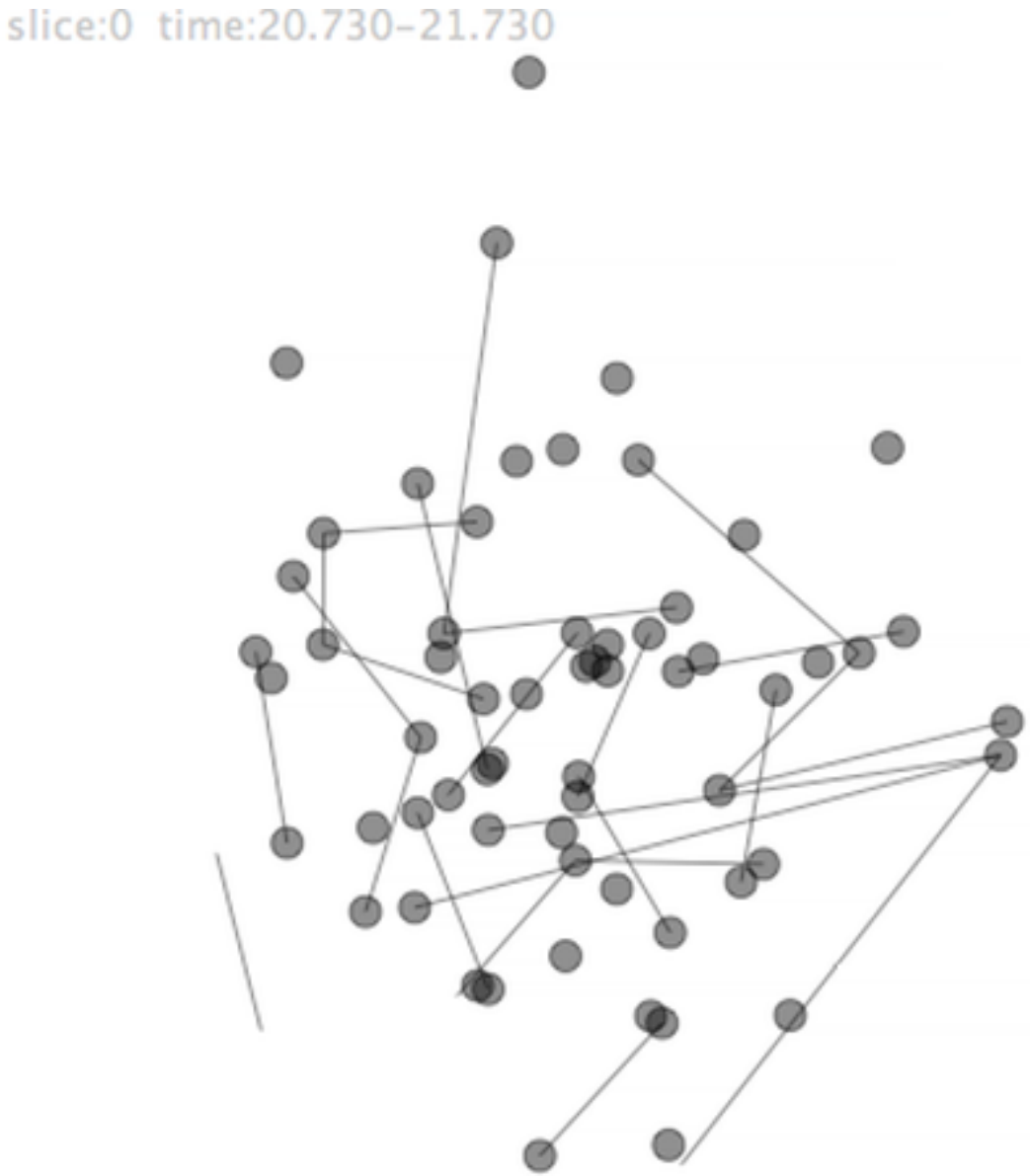


Mirco Musolesi
School of Computer Science
University of Birmingham

Joint work with Vito Latora, Cecilia Mascolo, Vincenzo Nicosia, Salvatore Scellato and John Tang

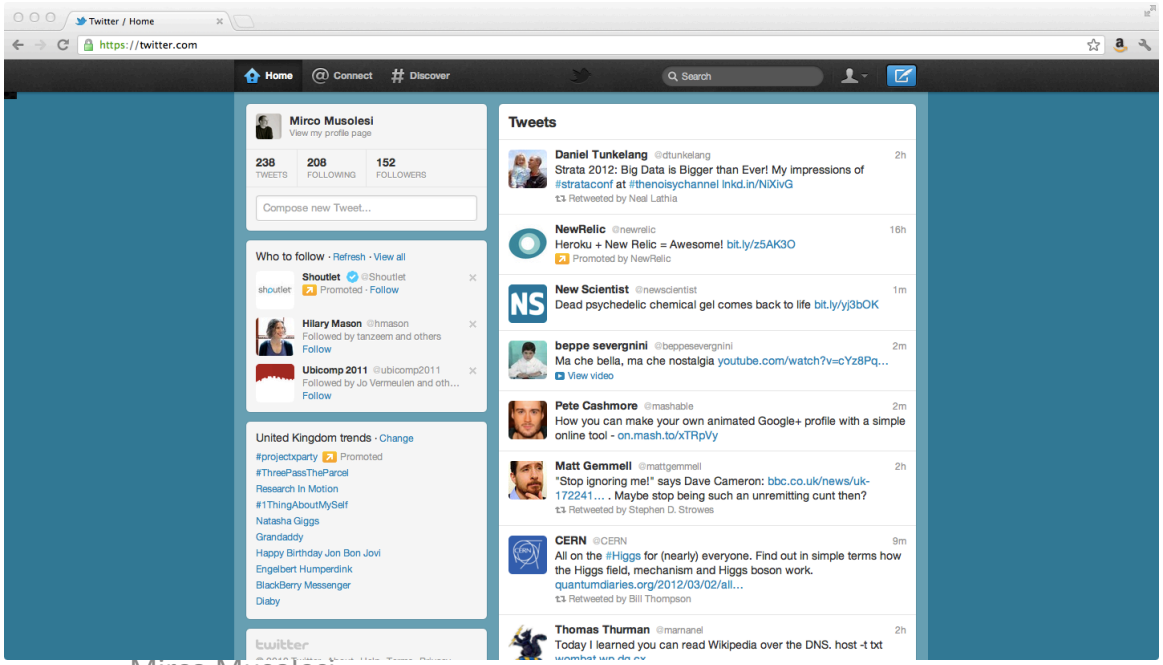


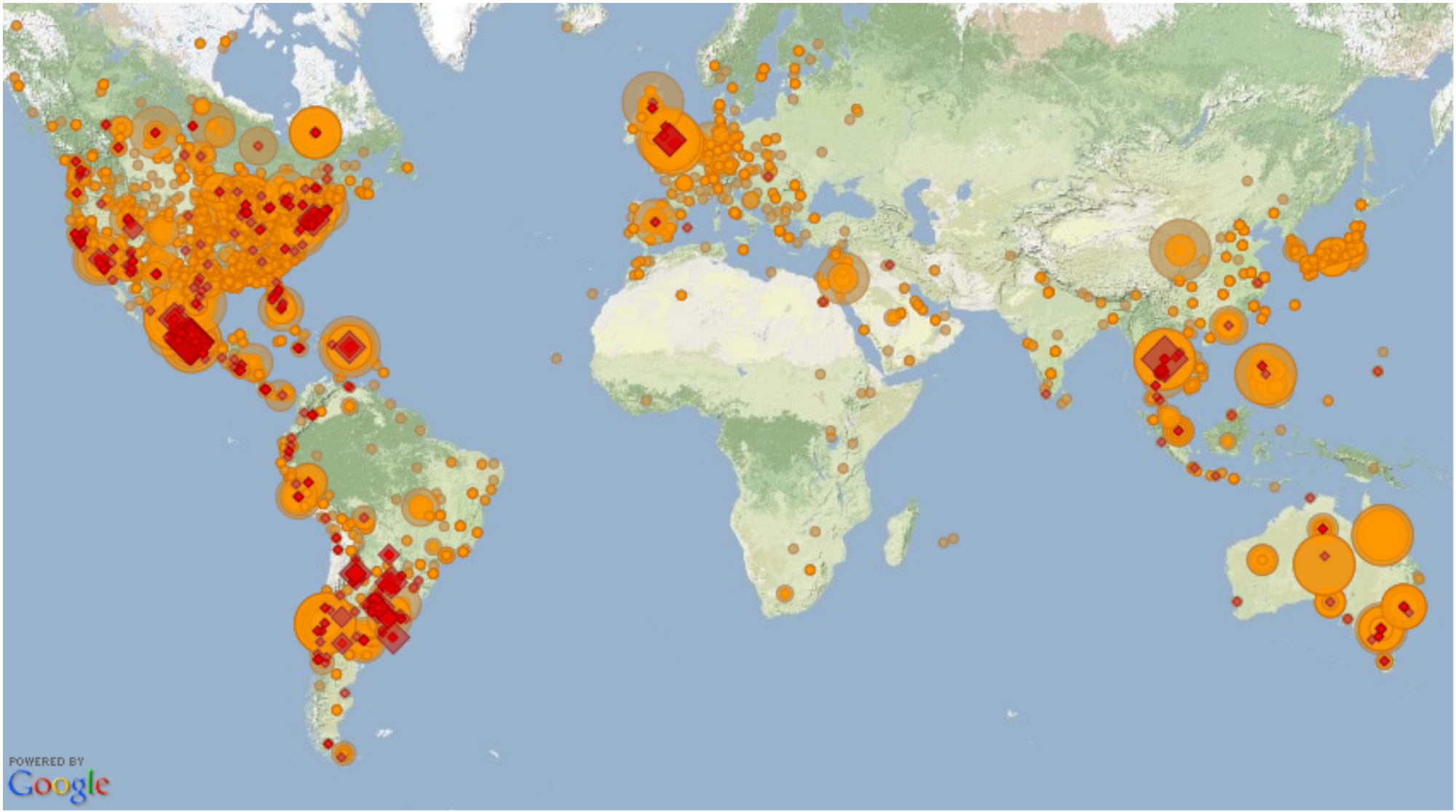
Credit: Mark Newman



Reality Mining Dataset

Facebook
https://www.facebook.com
facebook
Search
Mirco Musolesi
Home
Update status
Add photo/video
Ask question
Jamie Anderson
BBQSoc has only had 2 events and is already in the top 15 biggest societies.
Like · Comment · 3 minutes ago
Dan Winch likes this.
Dan Winch I'm joining!
55 seconds ago · Like
Write a comment...
Gianmarco Walden Ferrari
Við Munum Gera Betur Næst
Þetta Er Ágætis Byrjun
We'll Do Better Next Time
but This Is a Good Start
Sigur Rós- Heima- Ágætis Byrjun- HD
www.youtube.com
Fragmento del documental realizado por los integrantes de la banda islandesa Sigur Rós HEIMA. hermosa fotografía y un gran montaje.
Like · Comment · Share · 11 minutes ago
Loris Marinetti likes this.
Write a comment...
Samuele Cassani
mi sento lievemente disturbato
Like · Comment · 6 minutes ago near Bologna, Emilia-Romagna
Salvatore Scellato
just loved Andrew Bird - Sifters
Matteo Rizzo's birthday is today
Create event
Sponsored
See all
College Careers Advisers
eduvoice.co.uk
6th and College Careers Advisers needed to complete online survey for a University. £5 paid in 48 hours for 10 mins. Click to begin
Netflix - 1 month FREE
signup.netflix.com
Unlimited films and TV shows. 1 month FREE!
Farrow & Ball Showroom
Visit our new Leamington Spa showroom for the finest in paint, wallpaper and complimentary in-store colour consultancy
8,202 people like Farrow & Ball.
Dress to Impress
smithfieldcase.com
Let us dress you in style. Tell us your tastes and we'll deliver you a personalised case to try at home. Pick up and delivery free.
Tiziana Degli Esposti
listened to Suite No. 8 in G major, HWV 441: 6. Gavotta by Evgeni Koroliov on AUPEO!
Ross Apted likes Andy Fowler's photo.
Jamie Anderson BBQSoc has only had 2 events and is already in the top 15 biggest societies.
Daniel Dui likes Deejay Libre's link.
Samuele Cassani mi sento lievemente disturbato
Gianmarco Walden Ferrari
Við Munum Gera Betur Næst
Þetta Er Ágætis Byrjun
We'll Do Better Next Time
but This Is a Good Start
Sigur Rós- Heima- Ágætis...
Tiziana Degli Esposti
Antonio Lima
Greg Bigwood
Hamed Haddadi
Iain Phillips
Ian Gent
Simone Ivan Conte
Tristan Henderson
Go online to see who's online to chat.
Search
FAVOURITES
News Feed
Messages
Events
APPS
Links
Apps and Games
LISTS
Close friends
Family
Cambridge Univer...
Dartmouth College
University of Birmingham
University of Cam...
University of St An...
UCL
University of Bolo...
Leamington Spa Area
GROUPS
Cambridge Networks Ne...
In ricordo di David...
Next Generation N...
The Computer Laboratory
Create group


Twitter / Home
https://twitter.com
Home
Connect
Discover
Search
Mirco Musolesi
View my profile page
238 TWEETS
208 FOLLOWING
152 FOLLOWERS
Compose new Tweet...
Who to follow · Refresh · View all
Shoutlet @Shoutlet
Promoted · Follow
Hilary Mason @hmason
Followed by tanzeem and others
Follow
Ubicomp 2011 @ubicomp2011
Followed by Jo Vermeulen and oth...
Follow
United Kingdom trends · Change
#projectxparty Promoted
#ThreePassTheParcel
Research In Motion
#1ThingAboutMySelf
Natasha Giggs
Grandaddy
Happy Birthday Jon Bon Jovi
Engelbert Humperdink
BlackBerry Messenger
Diaby
Tweets
Daniel Tunkelang @dtunkelang 2h
Strata 2012: Big Data is Bigger than Ever! My impressions of #strataconf at #thenoisychannel lnkd.in/NiXivG
Retweeted by Neal Lathia
NewRelic @newrelic 16h
Heroku + New Relic = Awesome! bit.ly/z5AK3O
Promoted by NewRelic
New Scientist @newscientist 1m
Dead psychedelic chemical gel comes back to life bit.ly/yj3bOK
beppe severgnini @beppesevergnini 2m
Ma che bella, ma che nostalgia youtube.com/watch?v=cYz8Pq...
View video
Pete Cashmore @mashable 2m
How you can make your own animated Google+ profile with a simple online tool - on.mash.to/xTRpVy
Matt Gemmell @mattgemmell 2h
"Stop ignoring me!" says Dave Cameron: bbc.co.uk/news/uk-172241... . Maybe stop being such an unremitting cunt then?
Retweeted by Stephen D. Strowes
CERN @CERN 9m
All on the #Higgs for (nearly) everyone. Find out in simple terms how the Higgs field, mechanism and Higgs boson work. quantumdiaries.org/2012/03/02/all...
Retweeted by Bill Thompson
Thomas Thurman @marnanel 2h
Today I learned you can read Wikipedia over the DNS. host -t txt

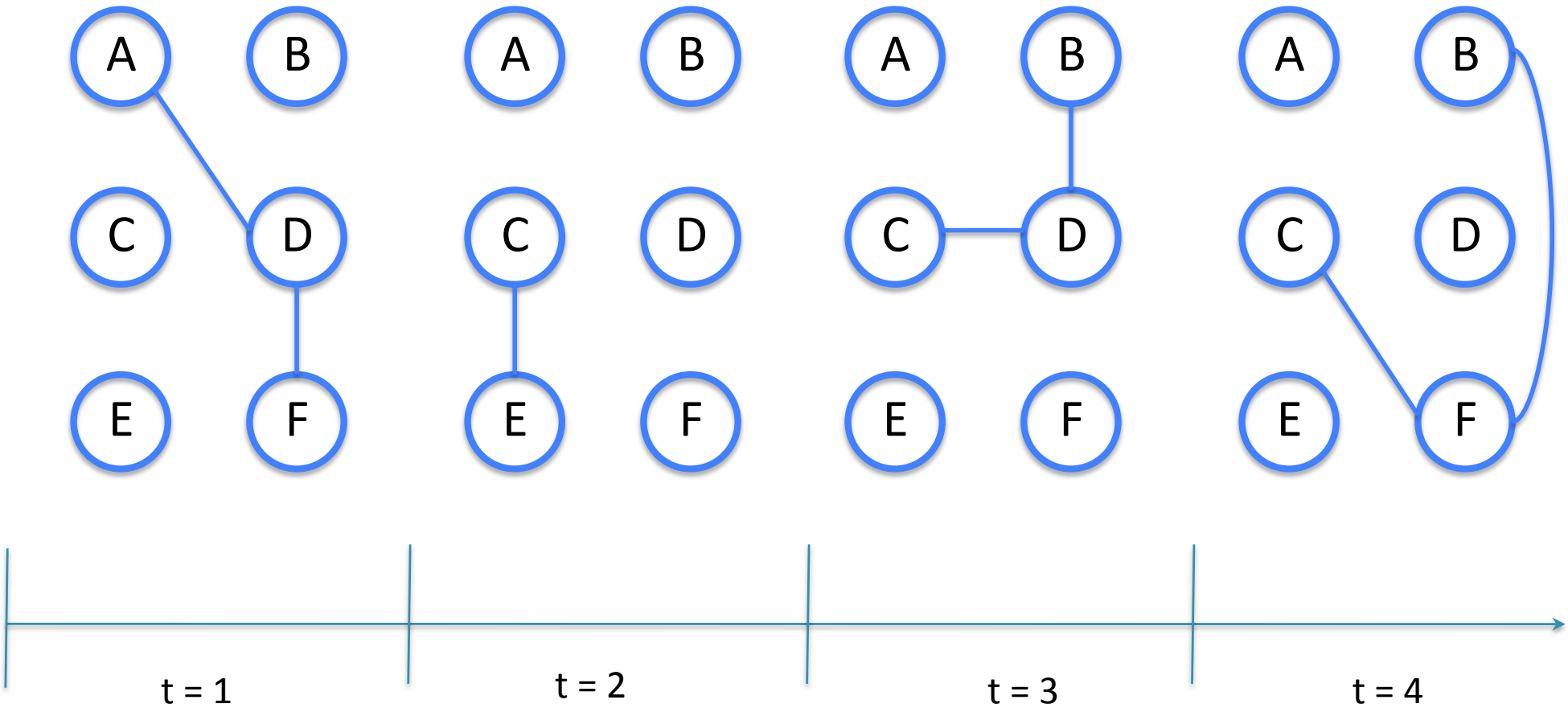


Credit: Flutracker.com

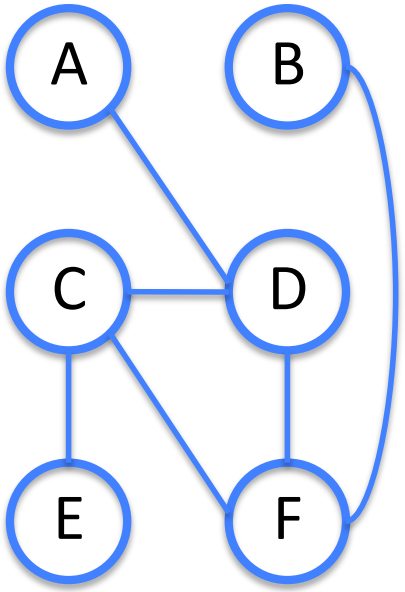
Problem: existing metrics do not capture the inherent dynamism of networks over time.

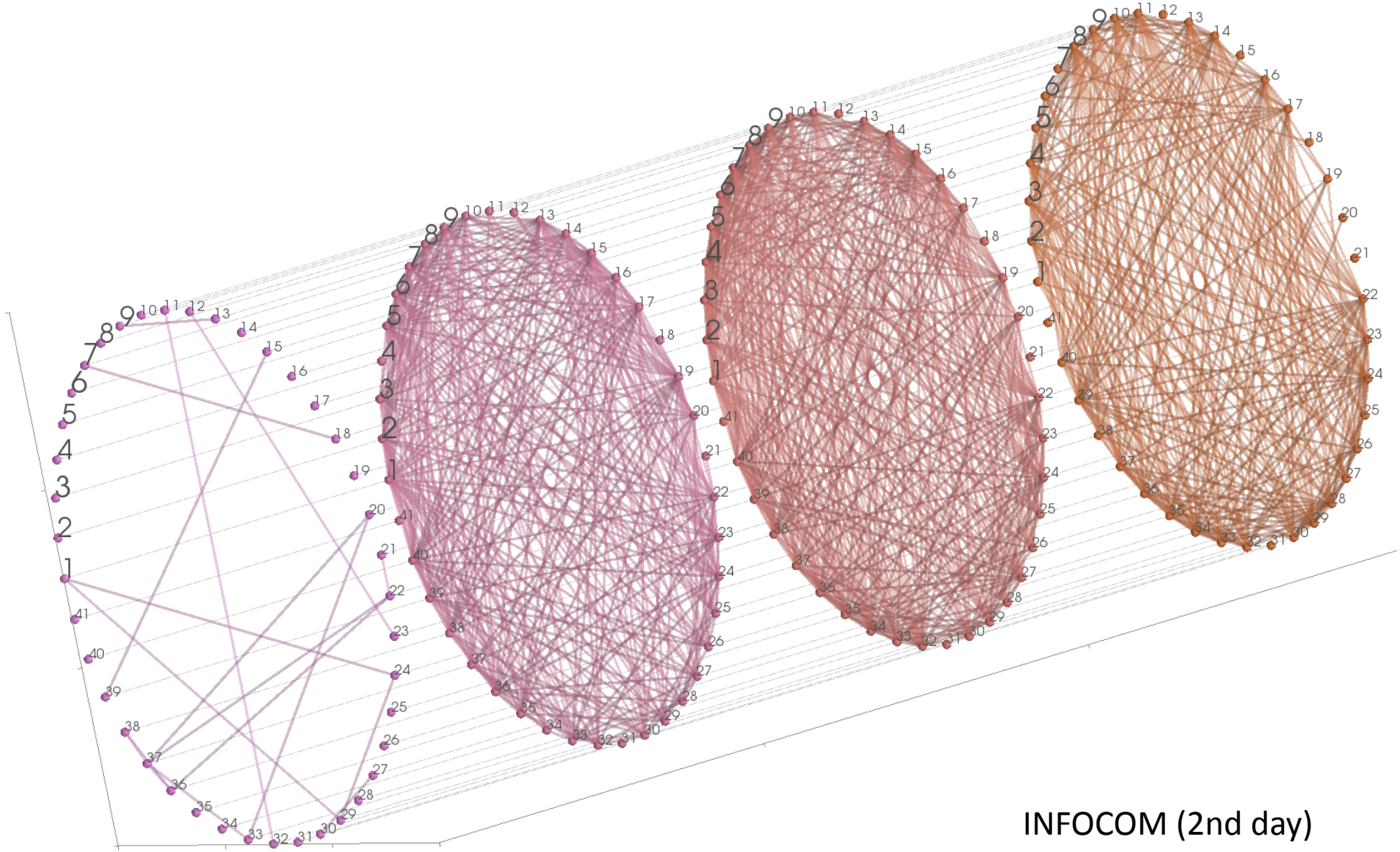
We need new **temporal metrics** defined over **temporal graphs** for studying dynamic processes over these networks.

An Example of Temporal Graph

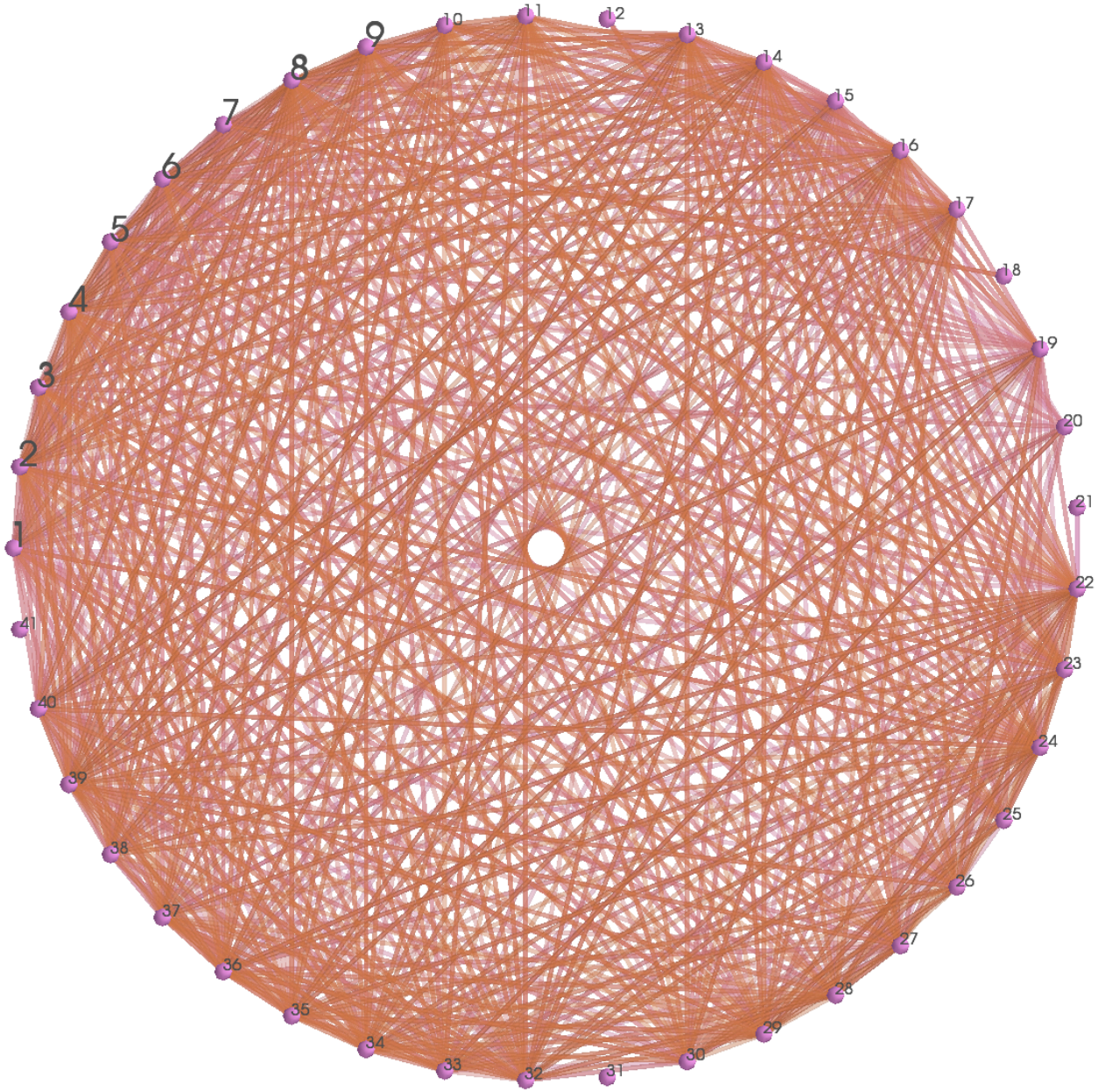


…and the Corresponding Static Graph



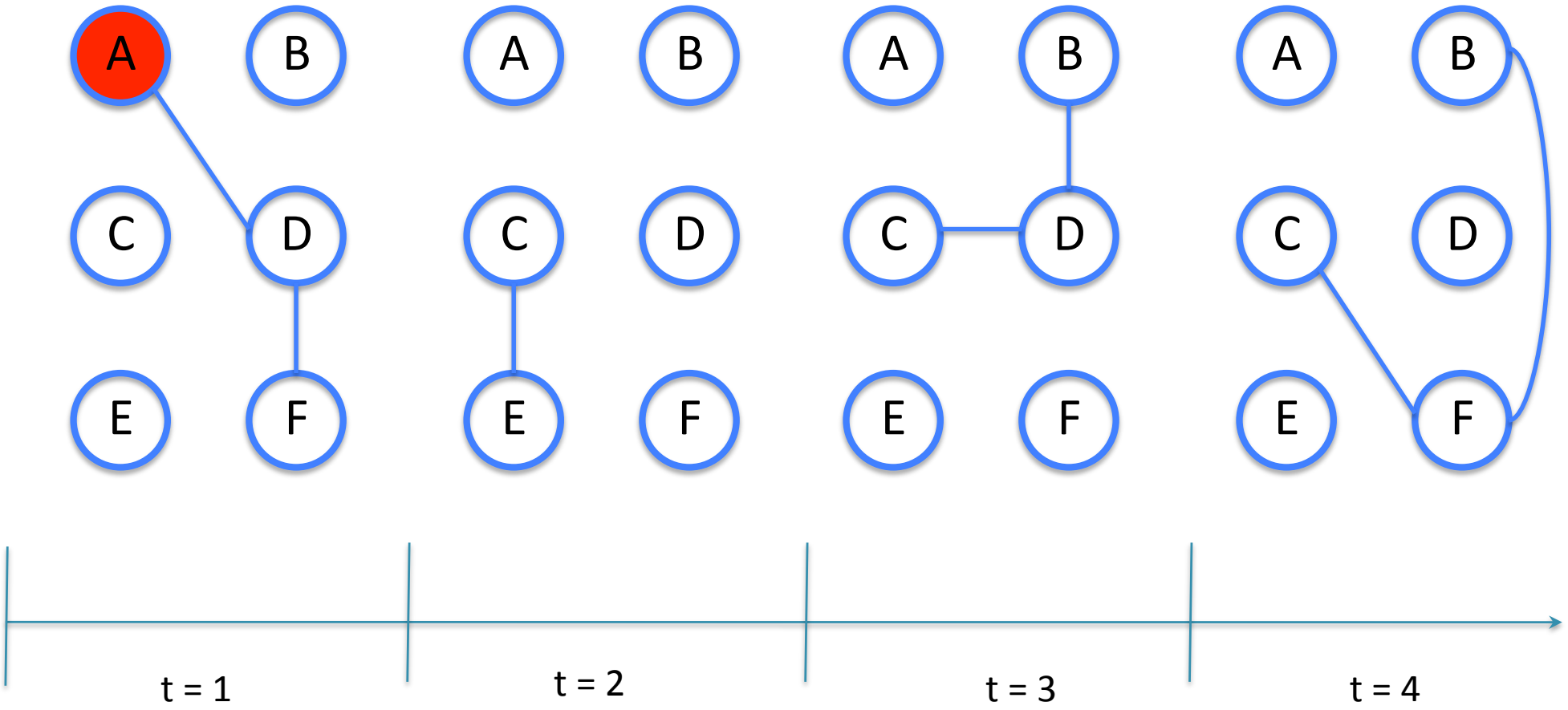


INFOCOM (2nd day)



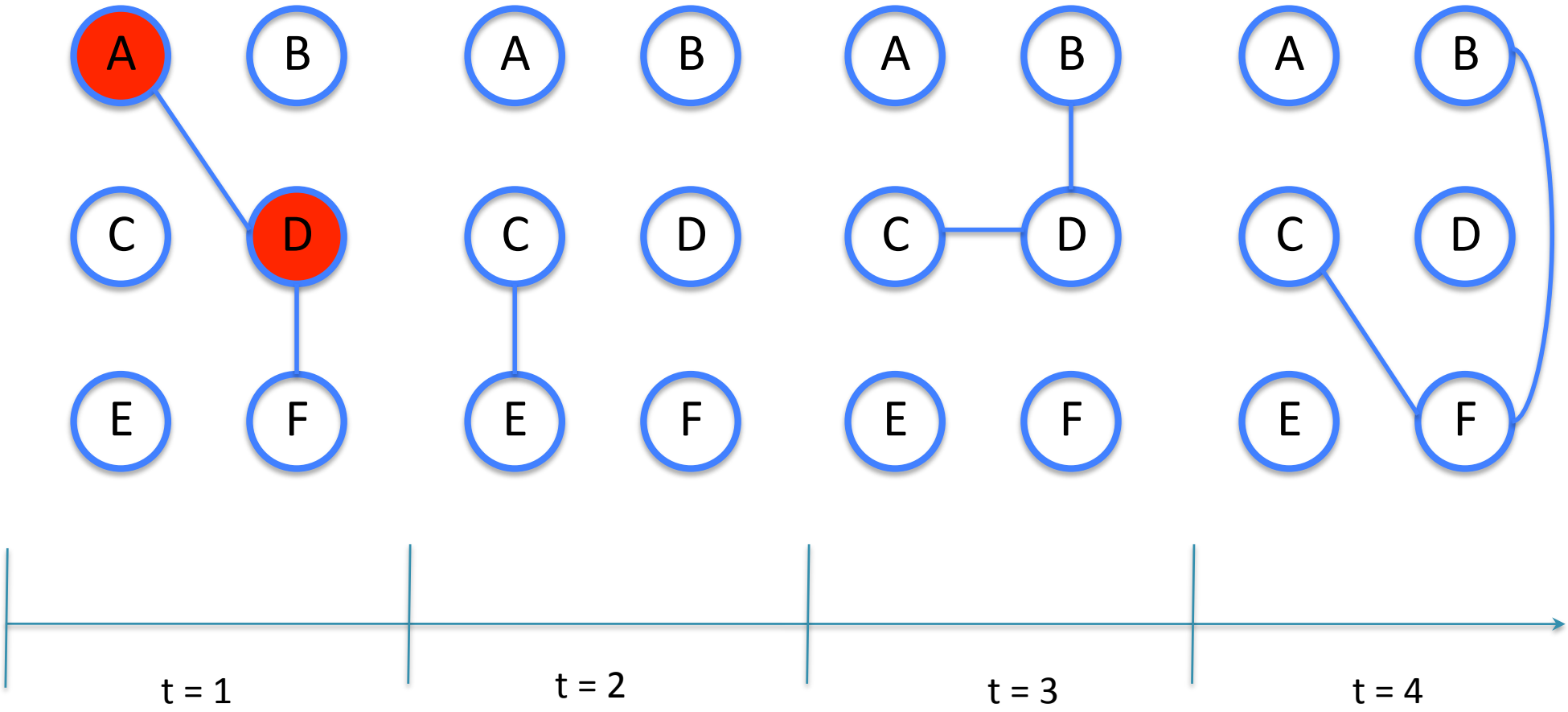
Mirco Musolesi

Calculating the Temporal Distance

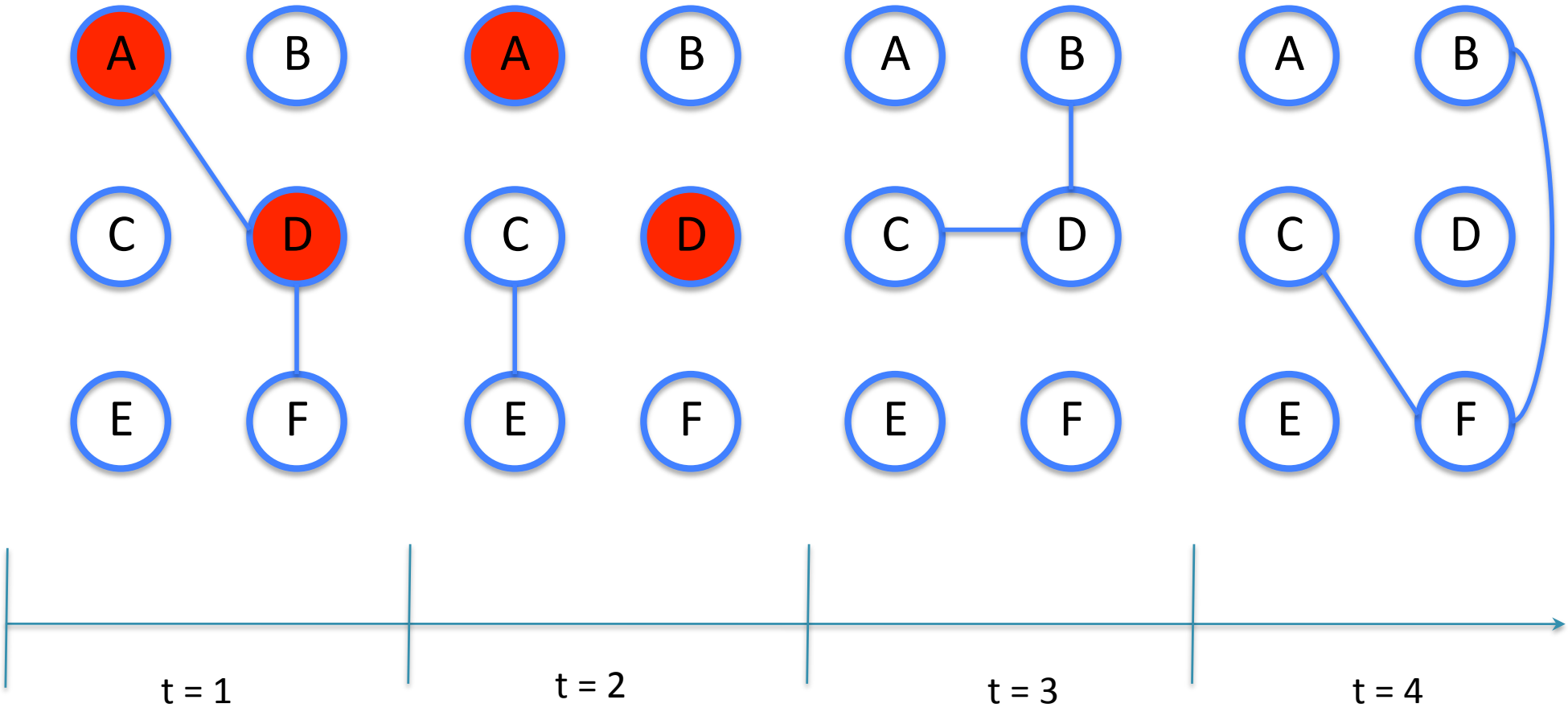


Calculating the Temporal Distance (t = 1)

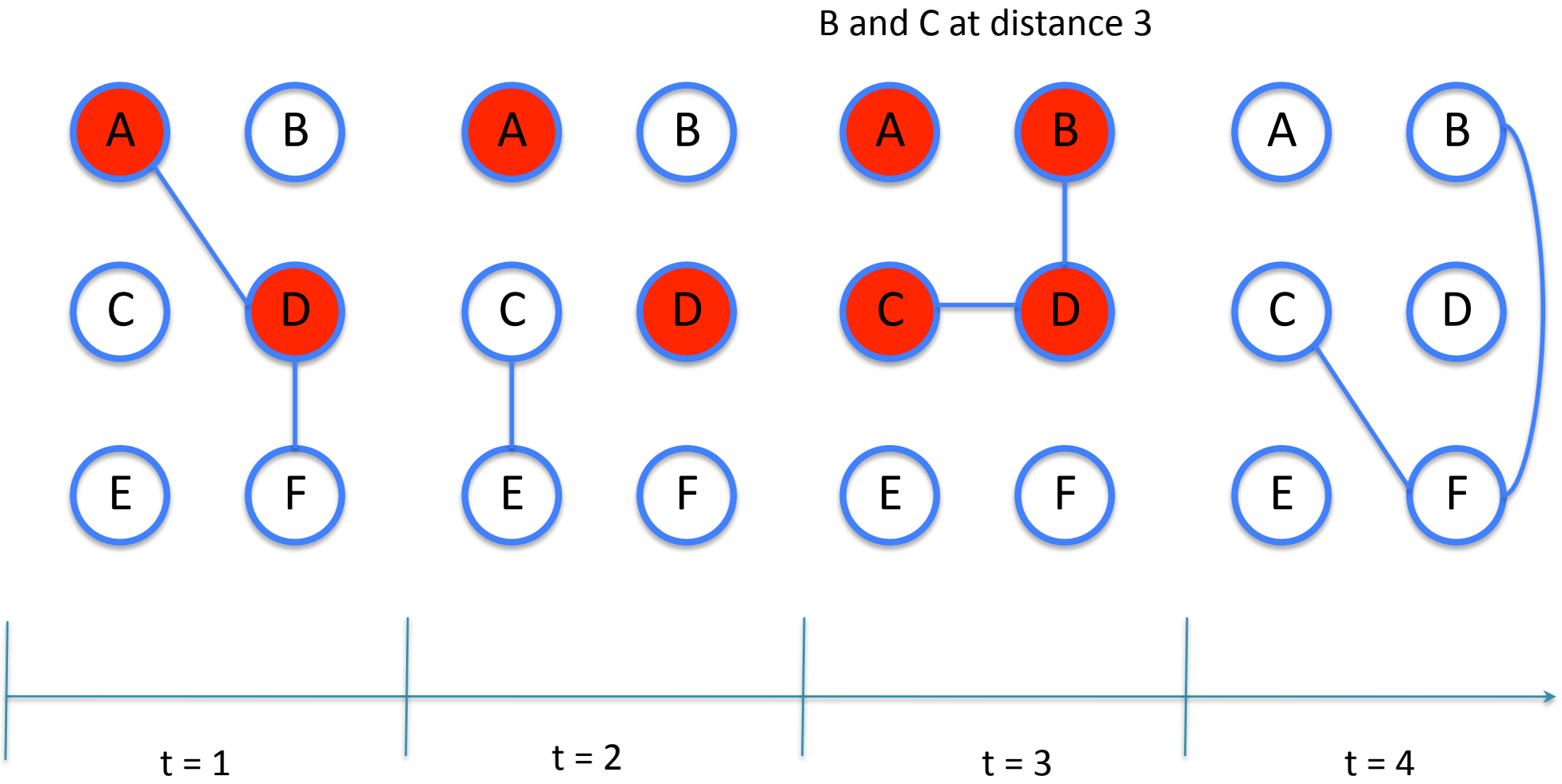
D at distance 1



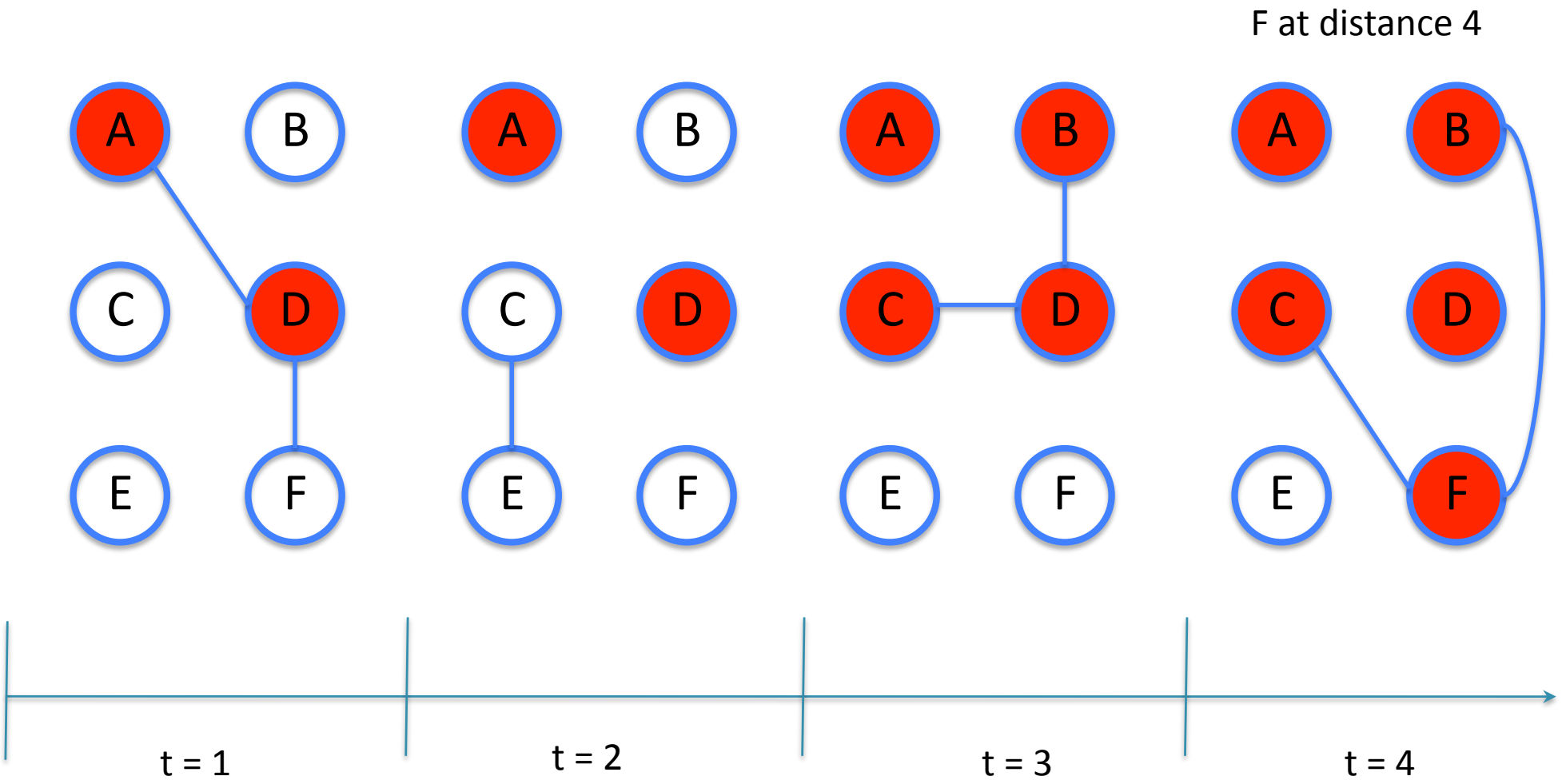
Calculating the Temporal Distance (t = 2)



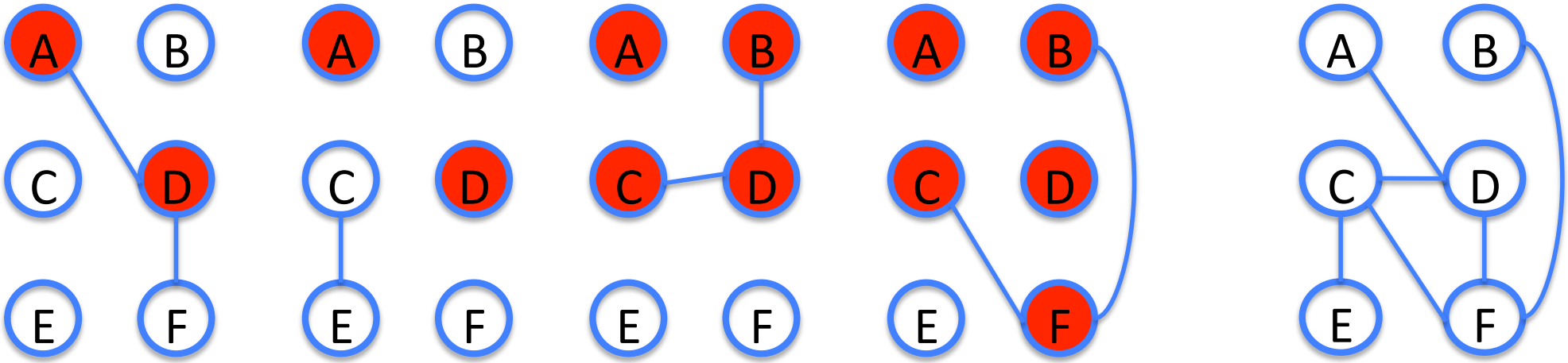
Calculating the Temporal Distance (t = 3)



Calculating the Temporal Distance (t = 4)



What about the Static Distance?

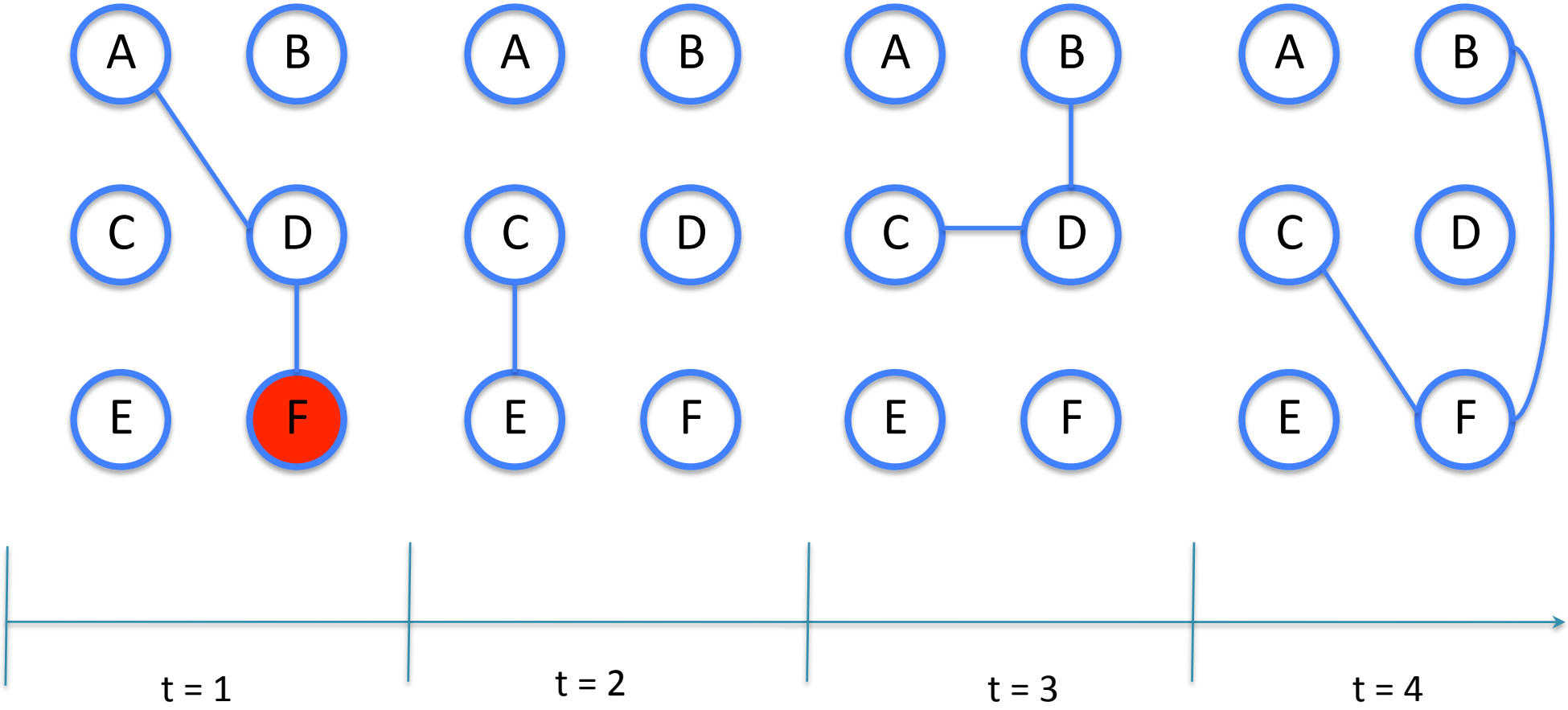


E is *statically* reachable but in reality it is not *dynamically* reachable!

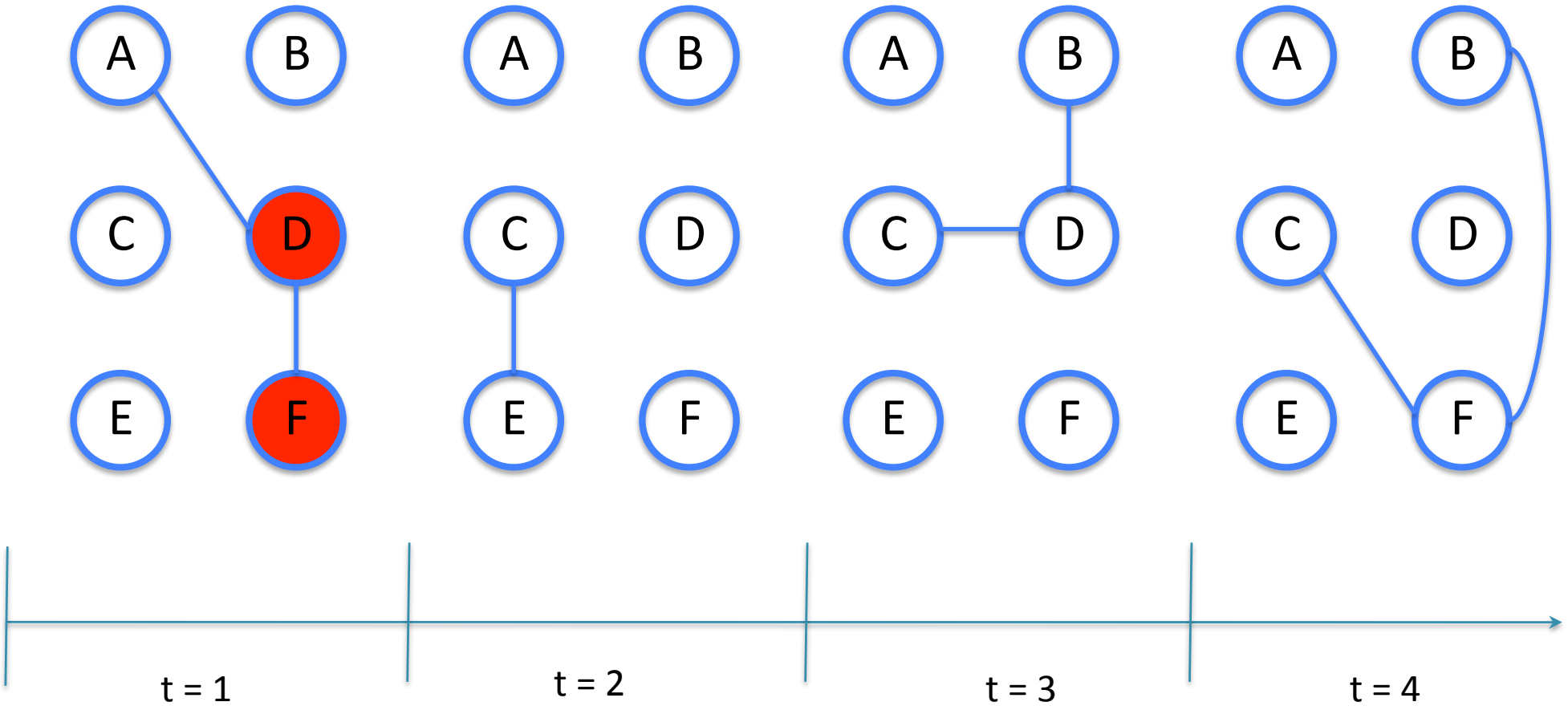
A-> F requires 2 transmissions (hops), but in reality it requires 3

No information about the duration of the process

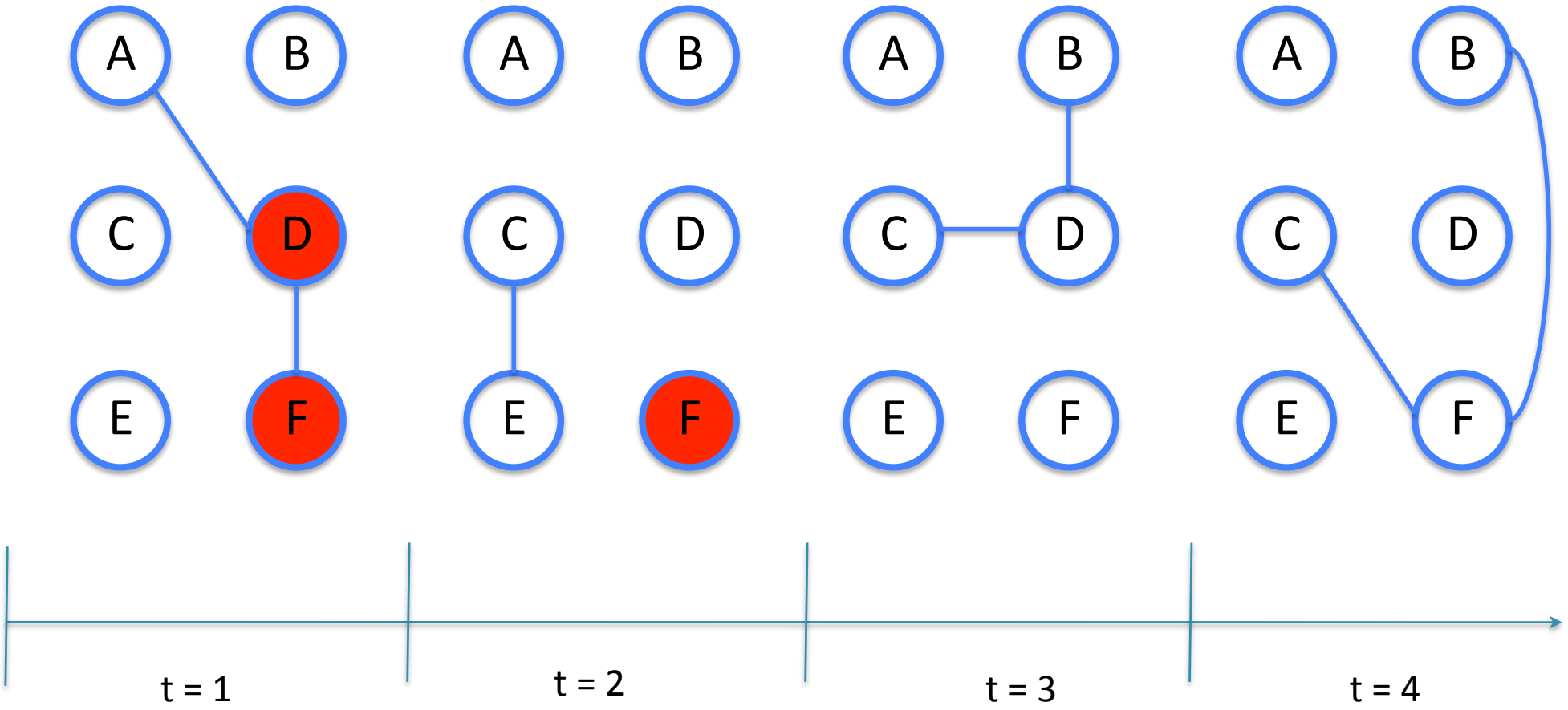
What about the Symmetric Distance (F to A)?



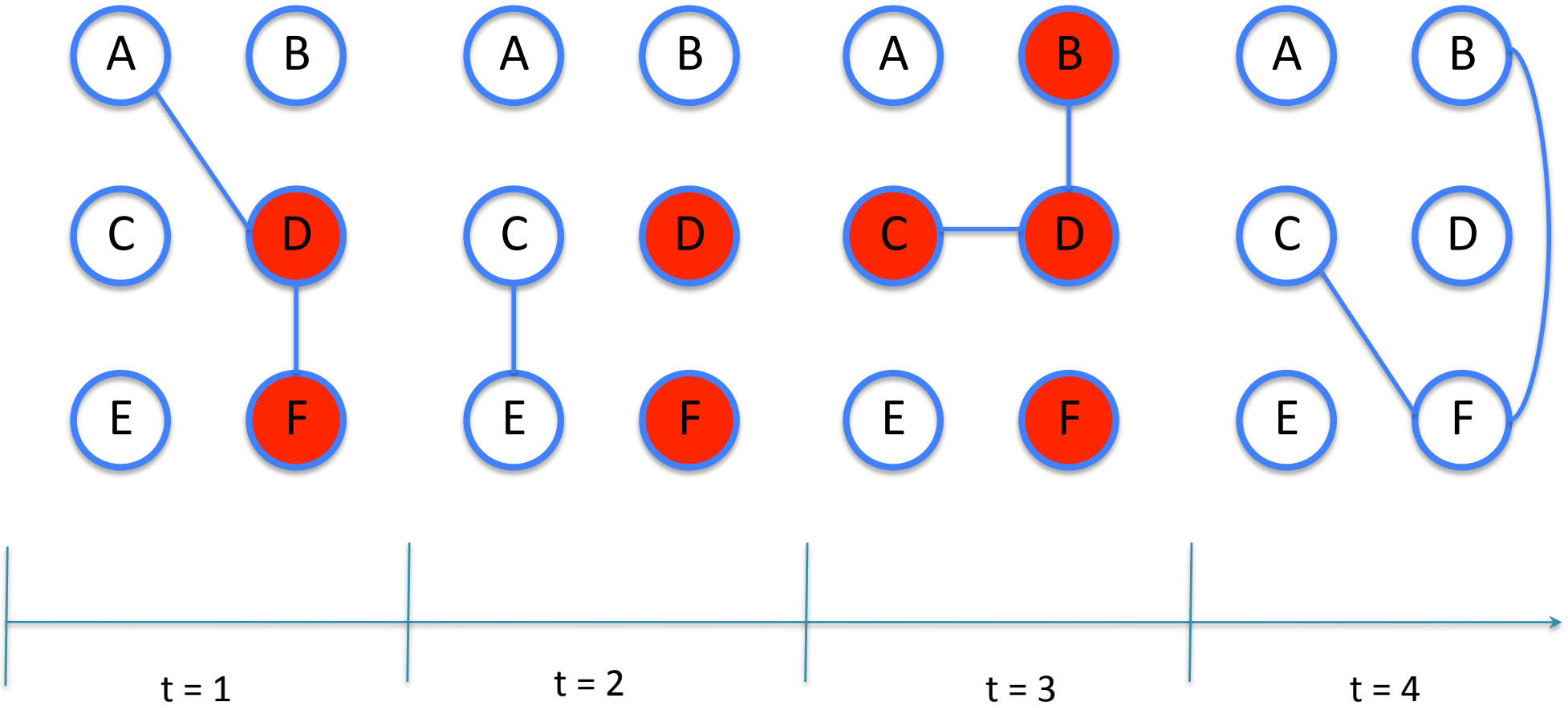
Calculating the Inverse Temporal Distance (t = 1)



Calculating the Inverse Temporal Distance (t = 2)

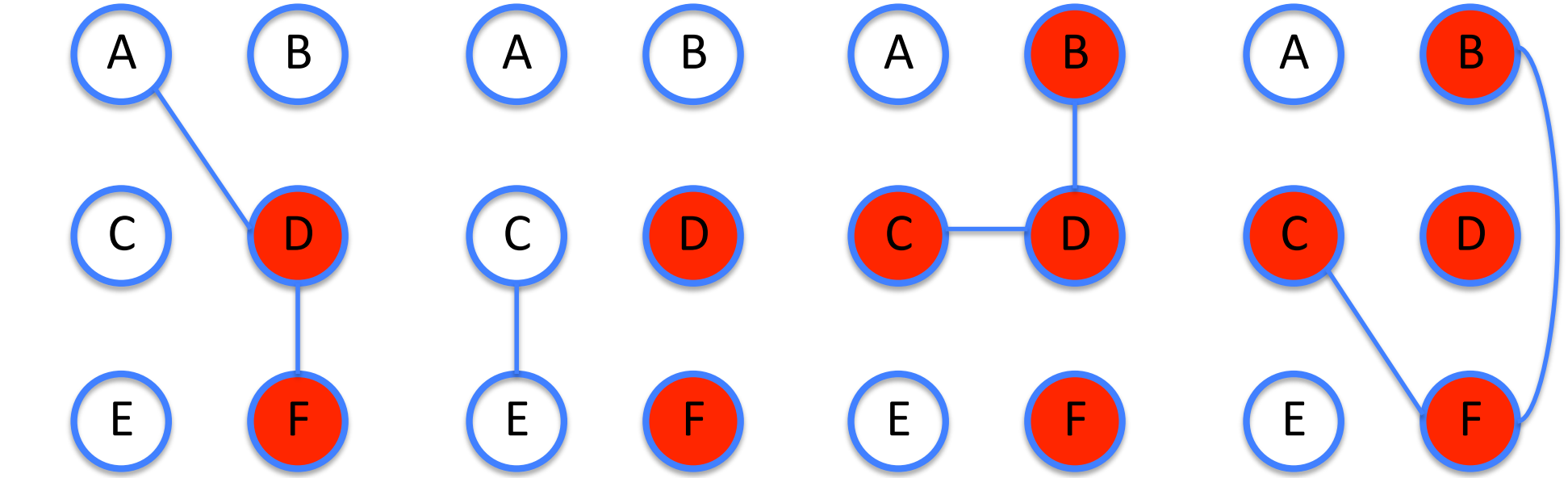


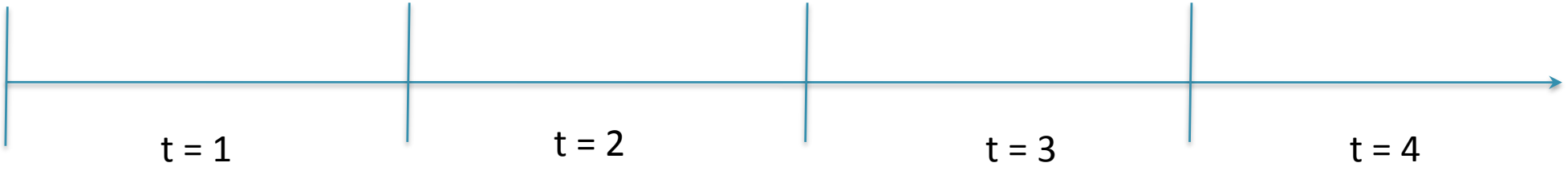
Calculating the Inverse Temporal Distance (t = 3)



Calculating the Inverse Temporal Distance (t = 4)

A is not reachable
[infinite distance]





Characteristic Temporal Path Length

- Characteristic temporal path length:

$$L^h(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} d_{ij}^h(t_{min}, t_{max})$$

- Defined considering the horizon of the infection
- Possible problem related to the potential divergence due to pairs of nodes that are not temporally connected

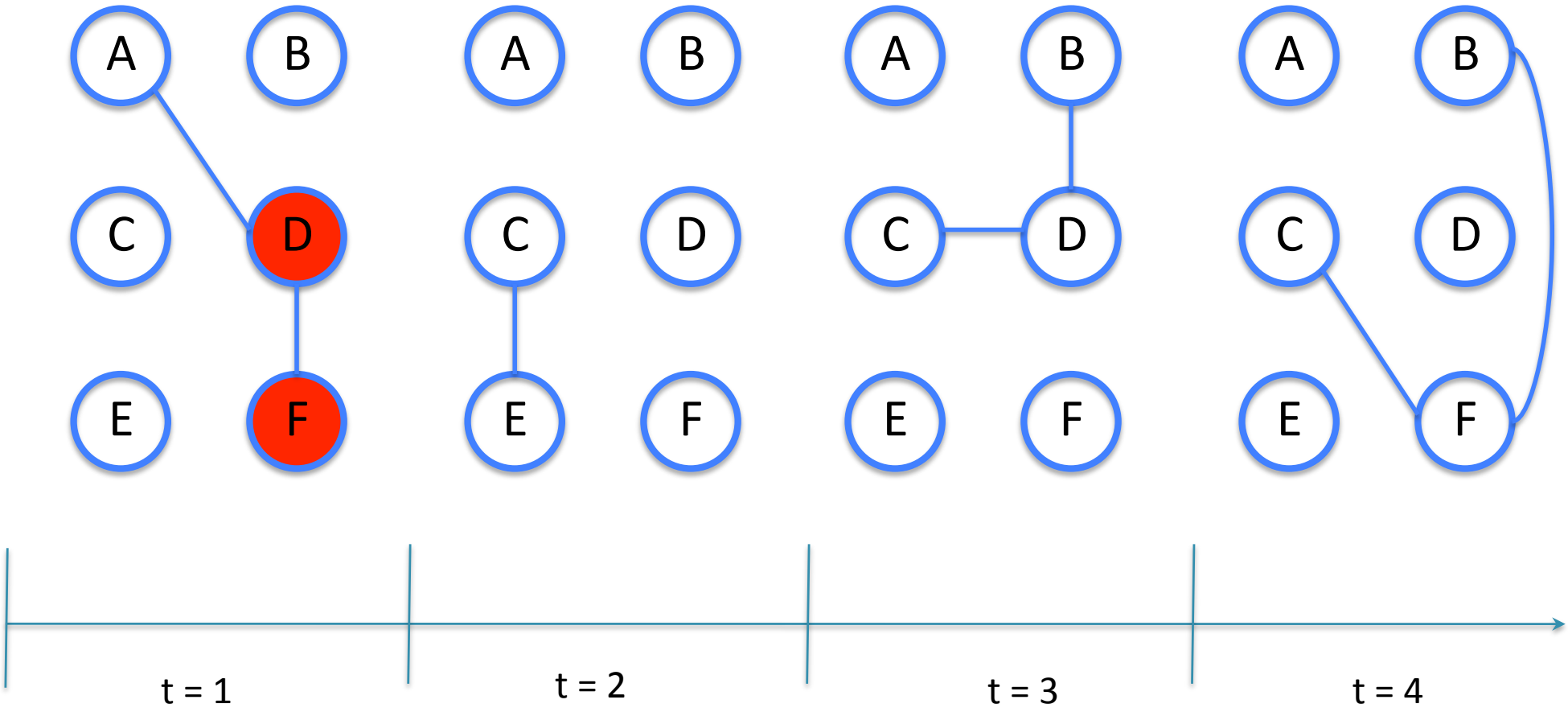
Characteristic Temporal Path Length

- Characteristic temporal path length:

$$L^h(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} d_{ij}^h(t_{min}, t_{max})$$

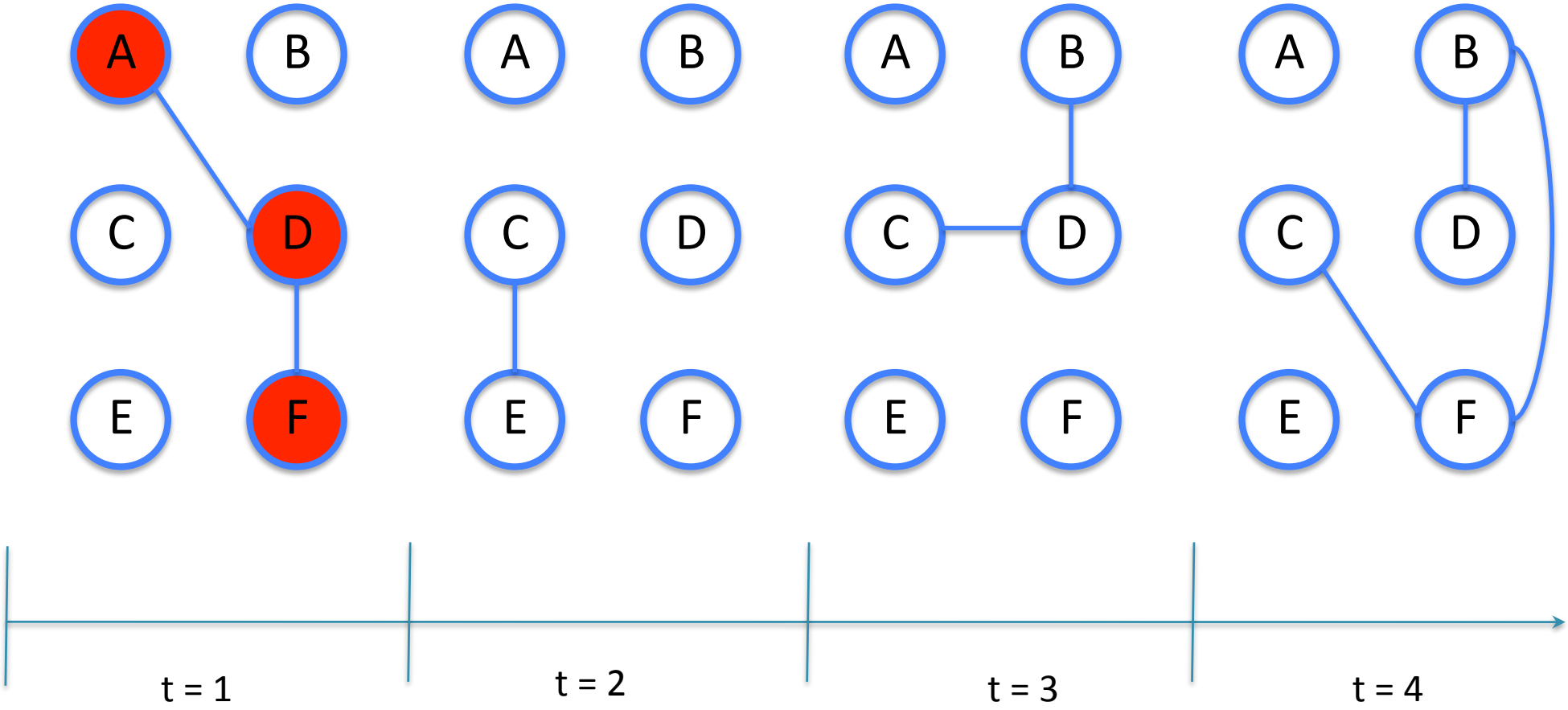
- **Defined considering the horizon of the infection**
- Possible problem due to the potential divergence due to pairs of nodes that are not temporally connected

Impact of the Horizon Parameter (F -> A, h = 1)



Impact of the Horizon Parameter (F -> A, h = 2)

A was not reachable at all with h = 1 (in 4 time windows), but with h = 2 it is a distance 1!



Characteristic Temporal Path Length

- Characteristic temporal path length:

$$L^h(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} d_{ij}^h(t_{min}, t_{max})$$

- Defined considering the horizon of the infection
- **Possible problem related to the potential divergence due to pairs of nodes that are not temporally connected**

Temporal Efficiency

- Solution: definition of temporal efficiency:

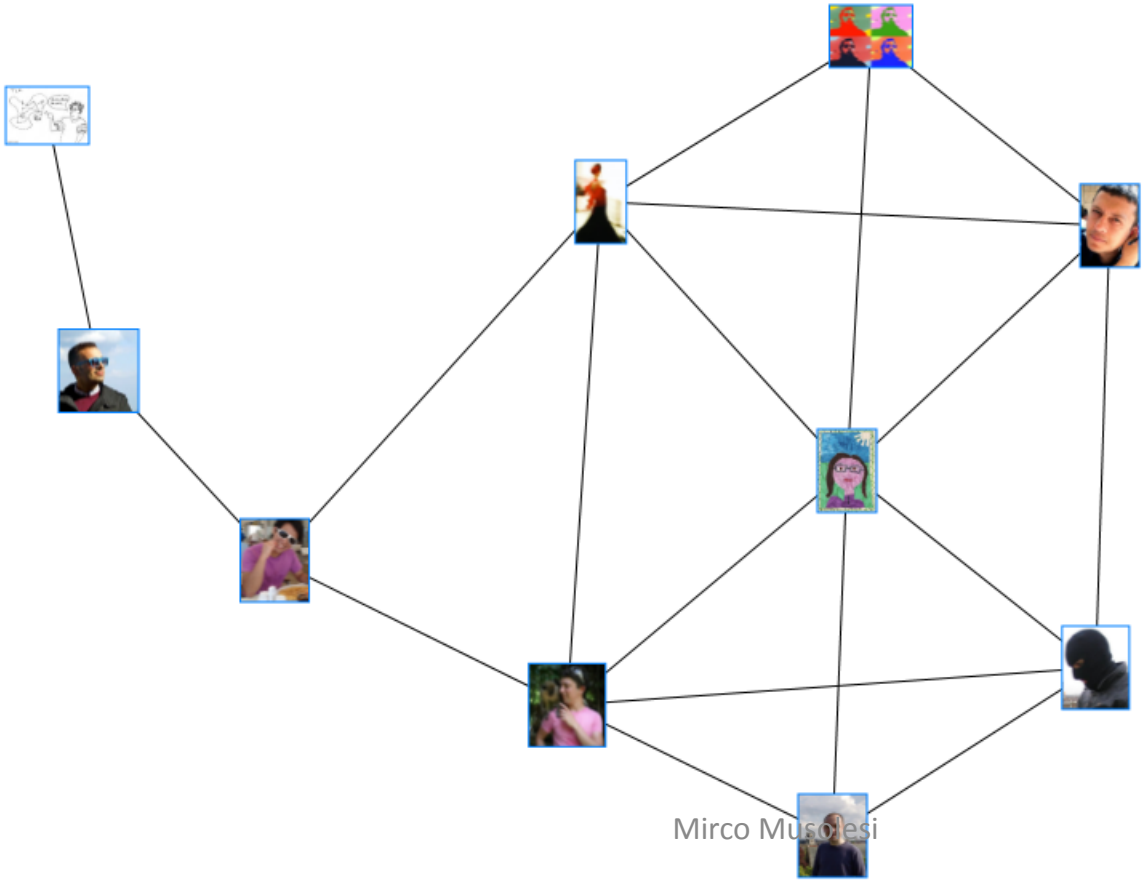
$$E^h_{T_{ij}}(t_{min}, t_{max}) = \frac{1}{d^h_{ij}(t_{min}, t_{max})}$$

$$E^h_{glob}(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} E^h_{T_{ij}}(t_{min}, t_{max})$$

- High value of E (low value of L) means that the nodes of the graphs can communicate efficiently

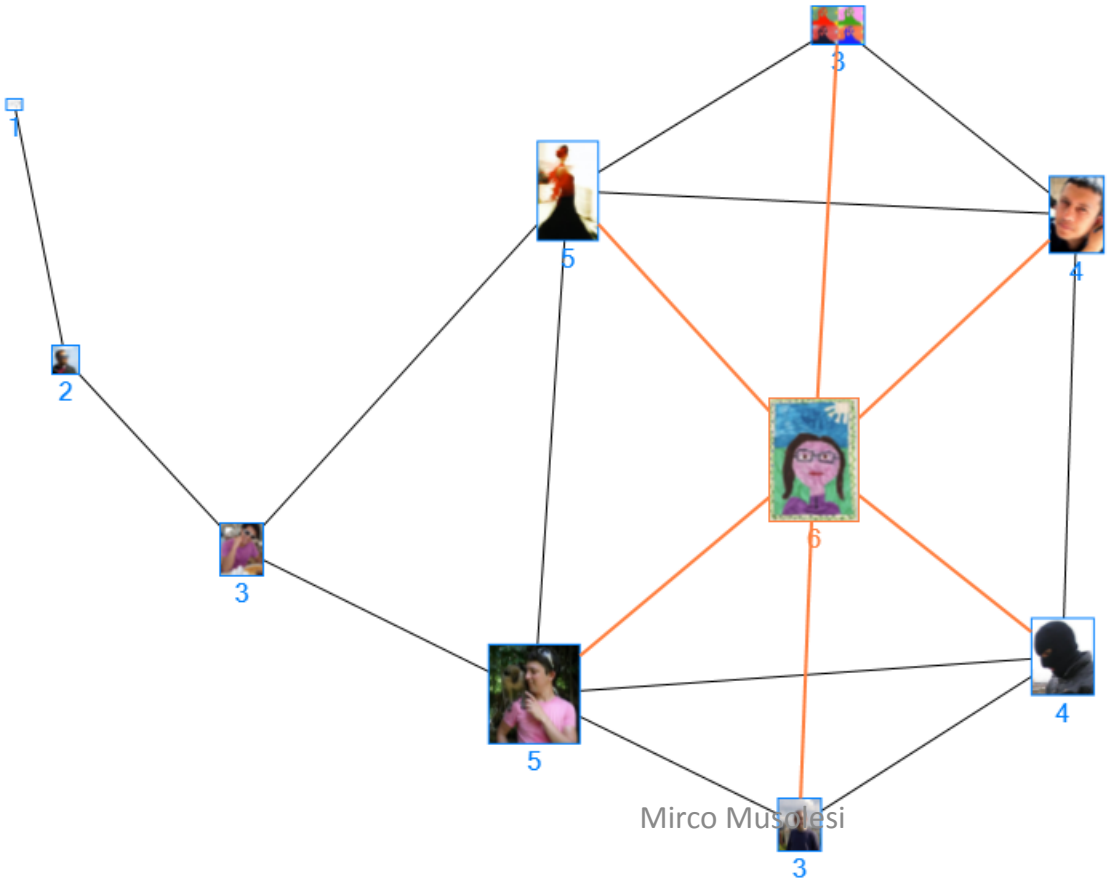
Centrality Metrics

- Most number of friends
- Quickly spread information to many people
- Mediates between the most information flows



Degree

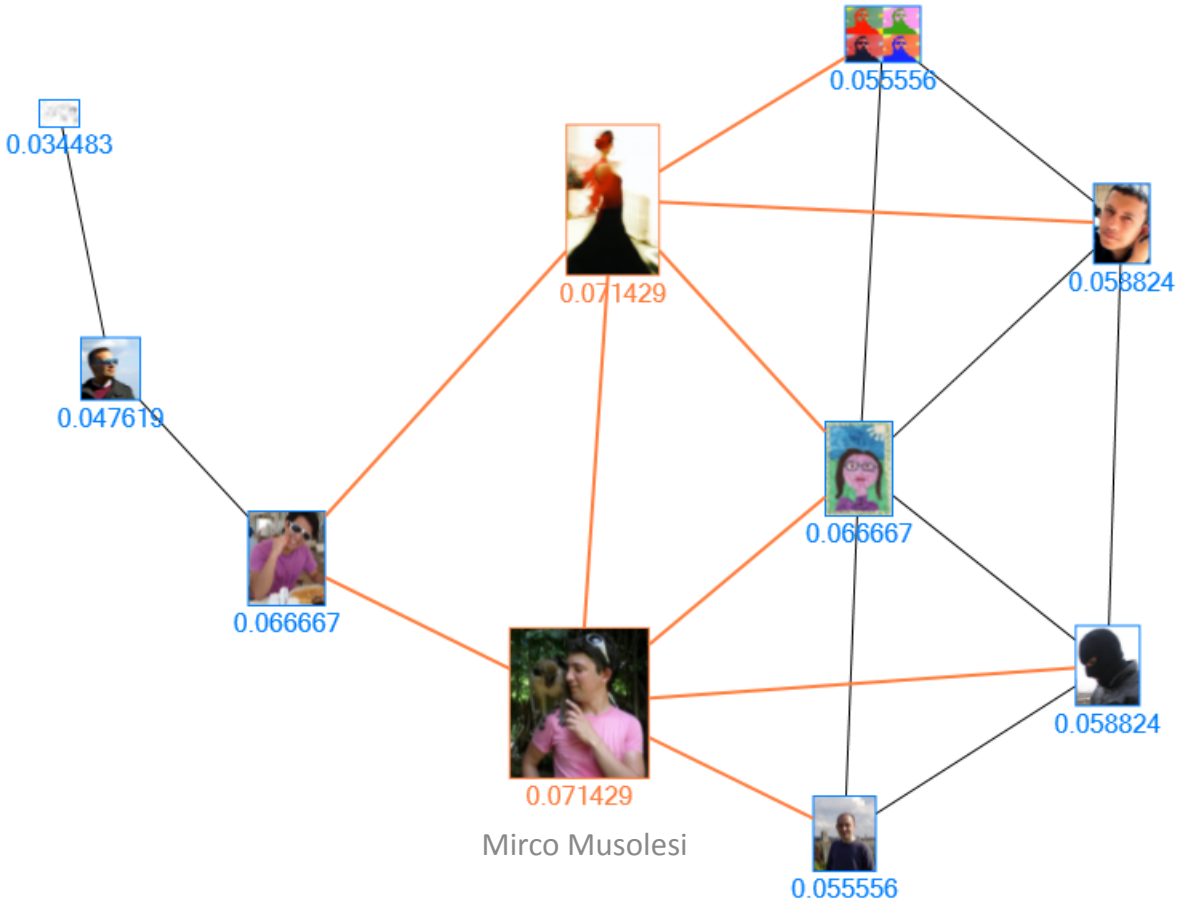
- $C_i^{deg} = \text{number of links to i}$
- Popular nodes



Mirco Musolesi

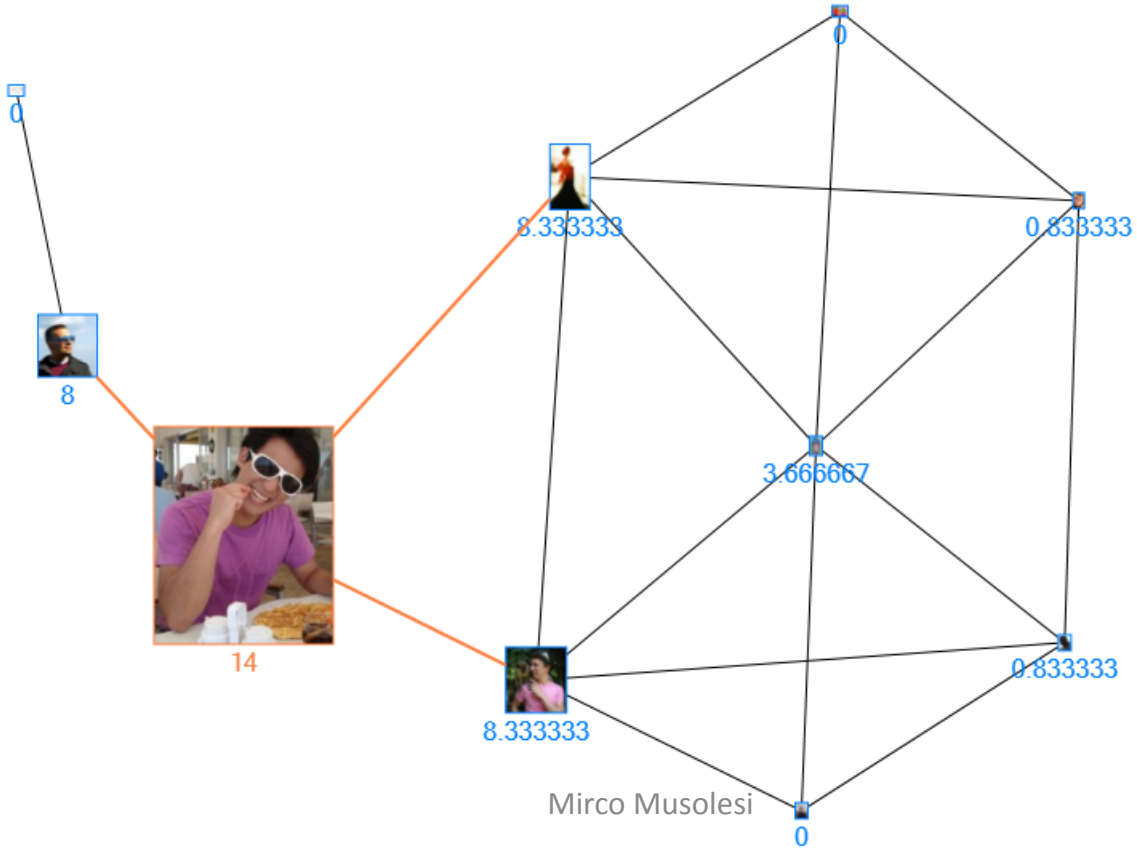
Static Closeness Centrality

- $C_i = \sum_{i \neq j} d_{ij}$
- Average shortest path length to all other nodes



Static Betweenness Centrality

- $C_i^{bet} = \sum_{i \neq s \neq t} \frac{\delta_s t(i)}{\delta_{st}}$ where δ_{st} is # shortest paths from s to t
  $\delta_{st}(i)$ is # shortest paths passing through i
- Fraction of shortest paths which pass through node i



Mirco Musolesi

Temporal Centrality Measures

- Static Closeness and Betweenness based on *static shortest paths*
- Definition of *closeness* and *betweenness* with temporal paths:
  - Duration
  - Time Order
  - Frequency

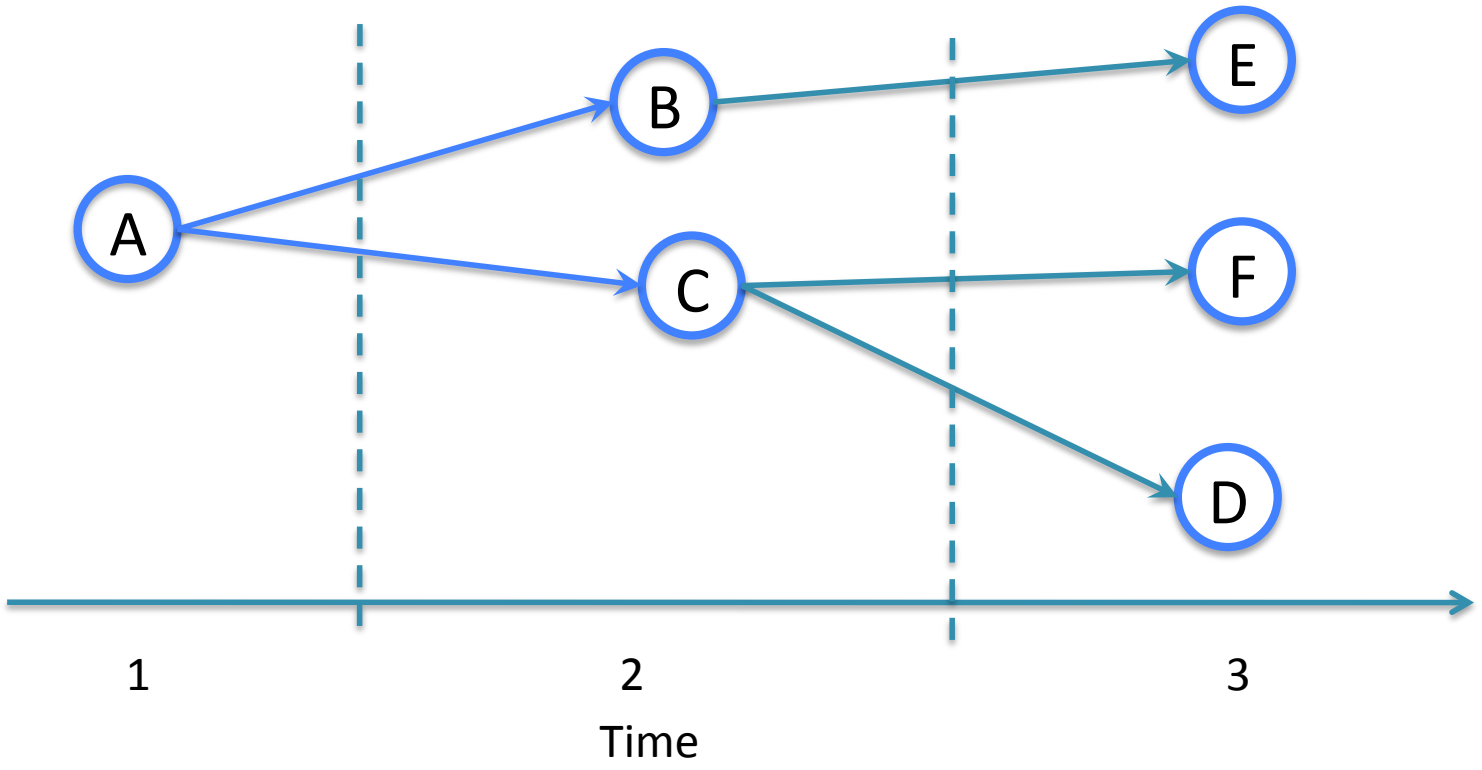
Temporal Closeness

Average over shortest *temporal* paths to all other nodes:

$$C_i = \frac{1}{W(N-1)} \sum_{j \neq i \in V} d_{i,j}$$

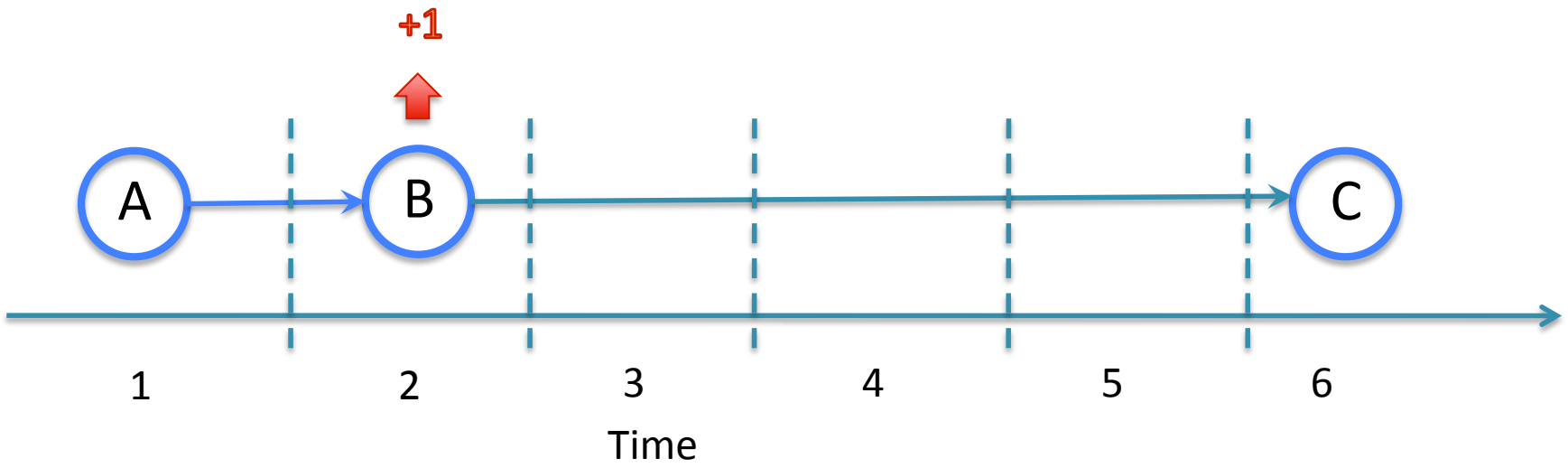
Temporal Closeness

$$C_A = \frac{(2+2)+(3+3+3)}{(3*(6-1))} = 0.867$$



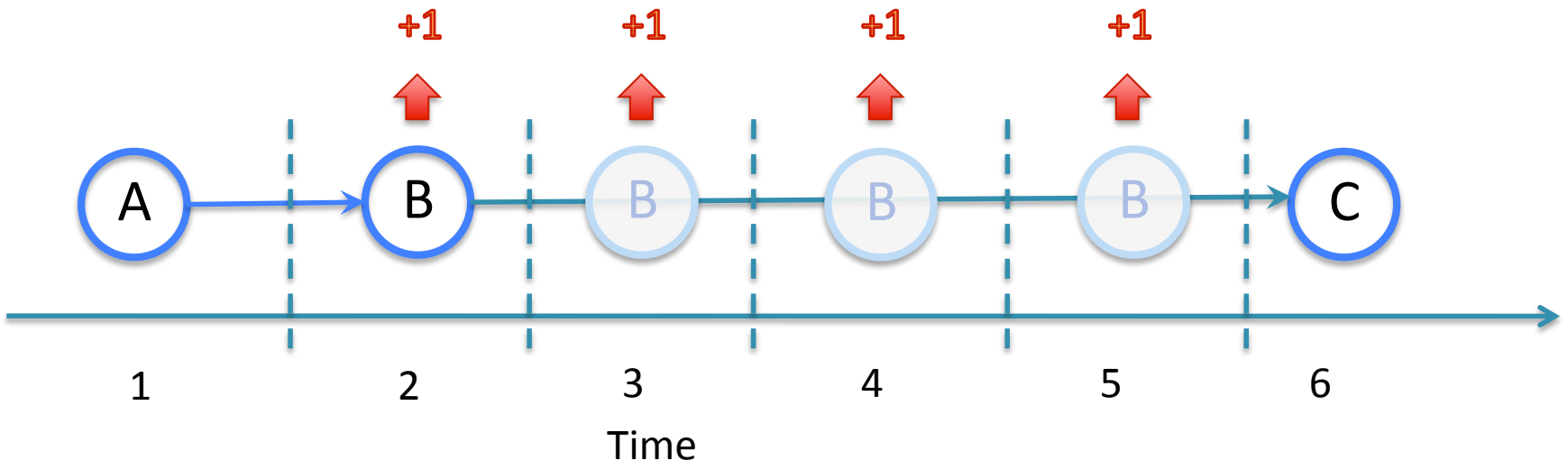
Temporal Betweenness

- Using temporal path length



Temporal Betweenness

- Take into account **duration**



Temporal Betweenness

$$C_i^B(t) = \frac{1}{(N-1)(N-2)} \sum_{\substack{j \in V \\ j \neq i}} \sum_{\substack{k \in V \\ k \neq i \\ k \neq j}} \frac{U(i,t,j,k)}{|S_{jk}^h|}$$

Where:

- $U(i,t,j,k)$ number of shortest paths from j to k, where node i is holding a message at time window t
- $|S_{jk}^h|$ number of shortest temporal paths between j and k

Temporal Betweenness

Sum over all time windows for each node:

$$C_i^B = \frac{1}{W} \sum_{t=1}^{W} C_i^B((t \times w) + t_{min})$$

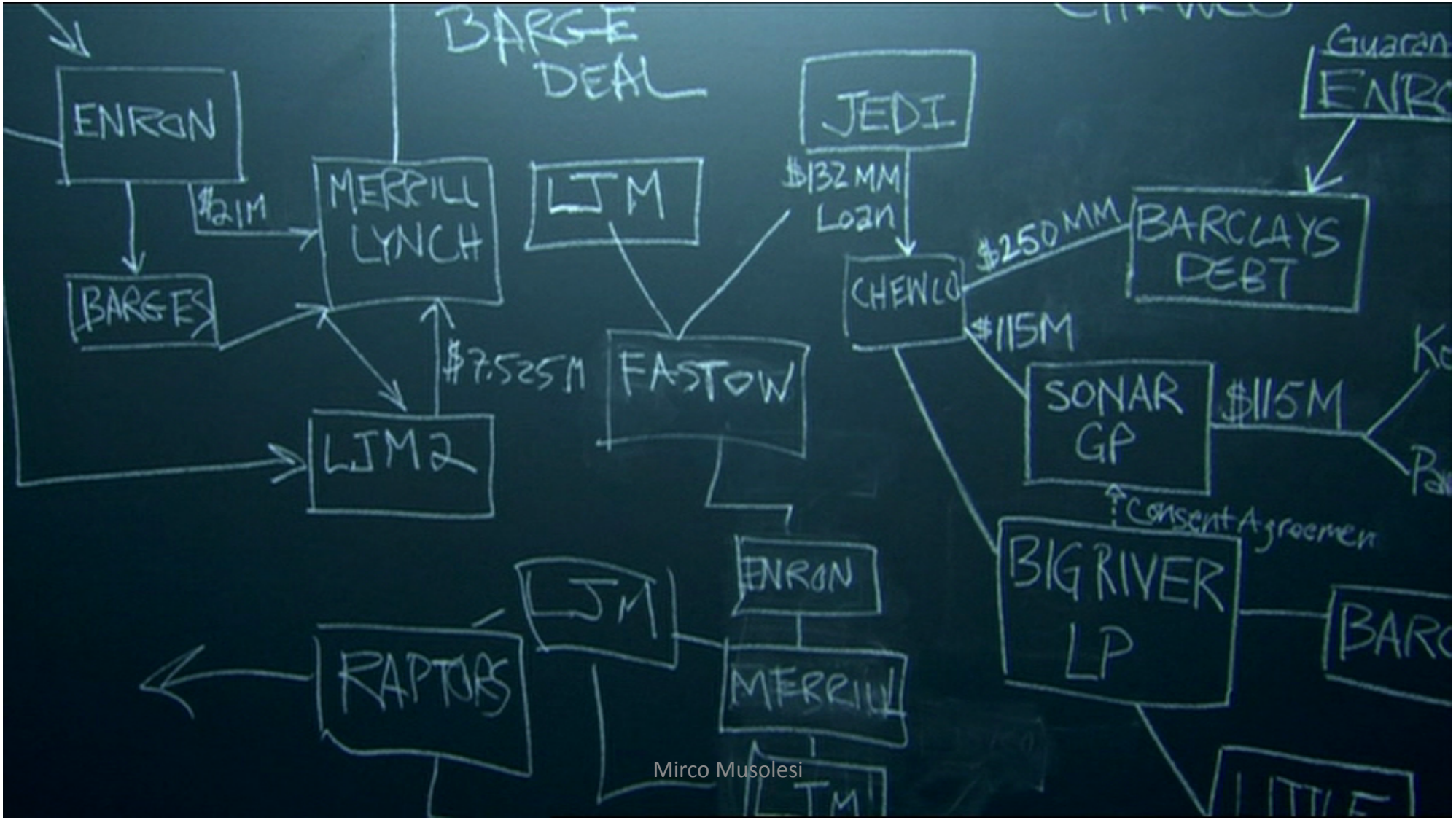
Evaluating Centrality

- Corporate Email Dataset
- Two perspectives:
  - Semantic: roles of each node
  - Dynamic Processes: simulate communication
    - Information Dissemination
    - Information Mediation

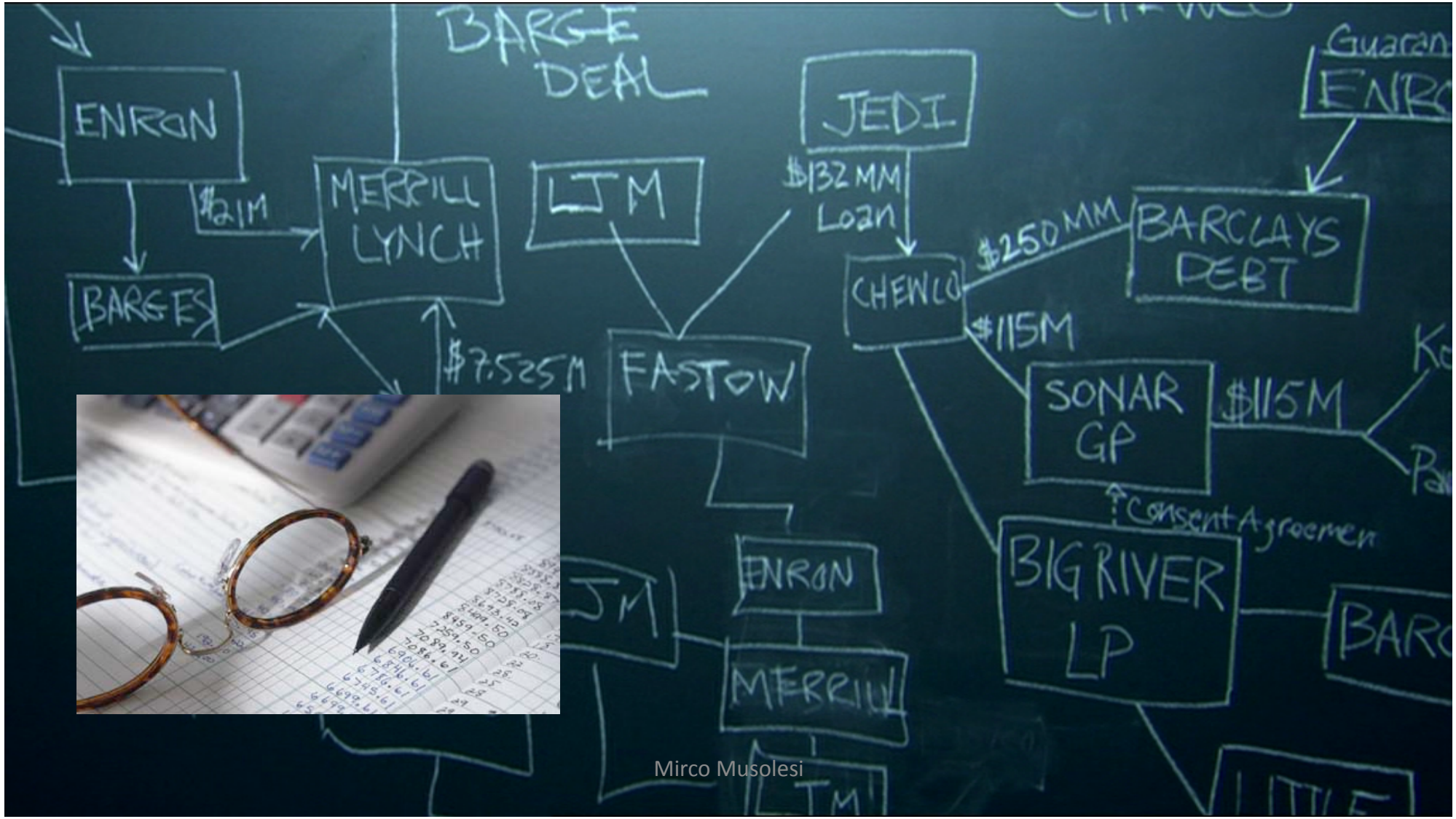
Evaluating Centrality: Enron in the News

Mirco Musolesi

Scandals


BARGE DEAL
ENRON
$12M
MERRILL LYNCH
LJM
JEDI
$132 MM Loan
BARGES
$7.525M
FASTOW
LJM2
CHEWCO
$250 MM
BARCLAYS DEBT
$115M
SONAR GP
$115M
Consent Agreemen
BIG RIVER LP
ENRON
JM
RAPTORS
MERRILL

Scandals


BARGE DEAL
ENRON
$21M
MERRILL LYNCH
BARGES
$7.525M
LJM
JEDI
$132 MM Loan
CHEWCO
$250 MM
BARCLAYS DEBT
$115M
FASTOW
SONAR GP
$115M
Consent Agreemen
BIG RIVER LP
ENRON
MERRILL
JM
LJM

Scandals

ENRON
BARGE DEAL
$12M
MERRILL LYNCH
LJM
BARGES
$7.525M
LIFE
NEWS
CELEBRITY
TRAVEL
ANIMALS
SPORTS
Return to Search Results
Civil Rights: Women in the Struggle
Taliban: Oppression and Resistance
Disasters at Sea
Troops Honor Fallen Comrades
EMAIL
RSS
RATE
PRINT
LINK TO
LICENSE
DIGG IT
TWEET
SHARE
California Experiences Second Week of Energy Crisis
384758 04: FILE PHOTO: Steam rises from the Etiwanda Generating Plant behind power lines near Rancho Cucamonga, CA, January 24, 2001, as the statewide energy crisis continues. The power plant is owned and operated by Edison O&M Service Division. Blackout warnings were issued for the first time by the California Independent System Operator June 17, reporting that blackouts could occure Monday and Tuesday afternoons between noon and eight in the evening.
Photo: David McNew/Getty Images
Jun 18, 2001

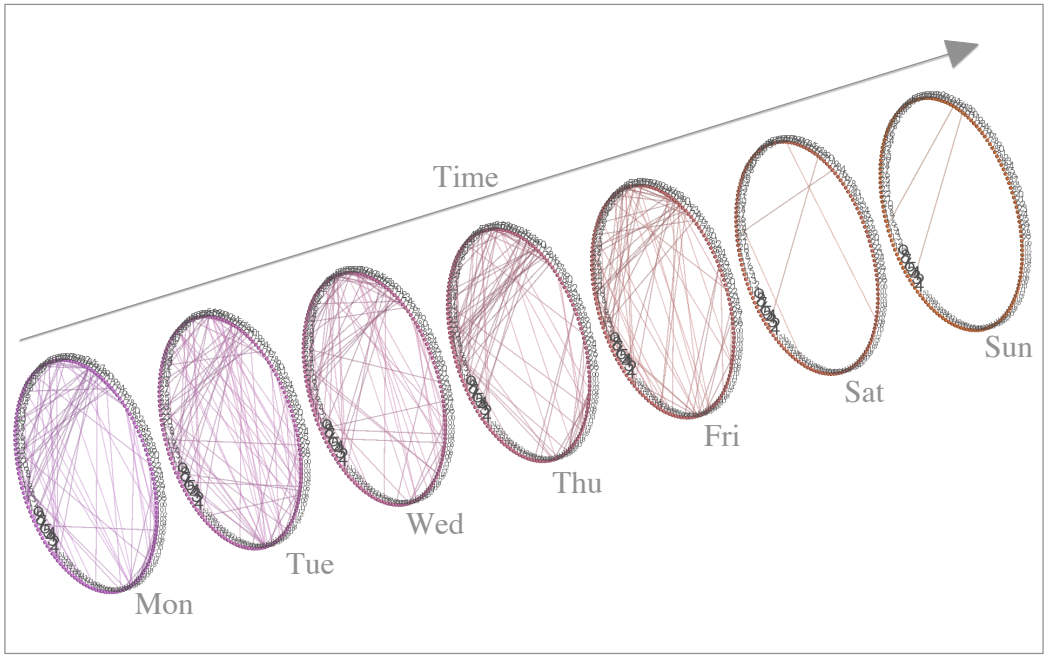
Public Investigation

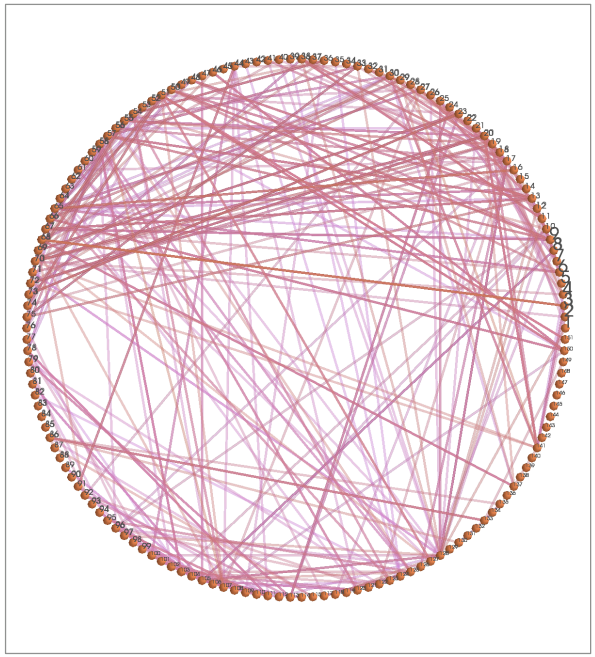
- Telephone logs
- Documents
- Financials
- Emails

- 151 user mailboxes
- May 1999 to Jun 2002
- 250,000 emails
- NOT anonymised

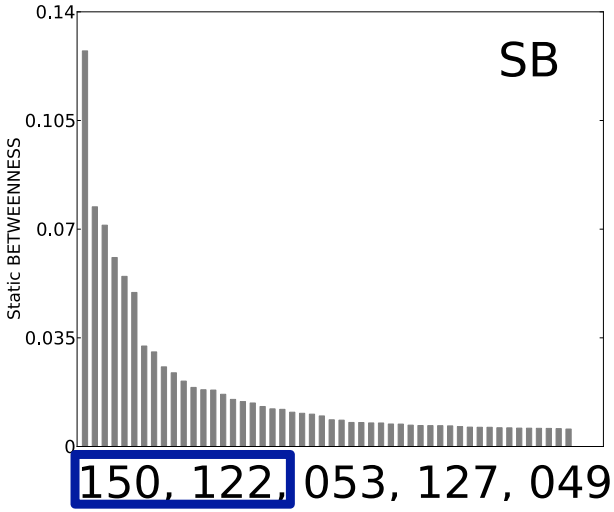
Email exchanges to Temporal Graph

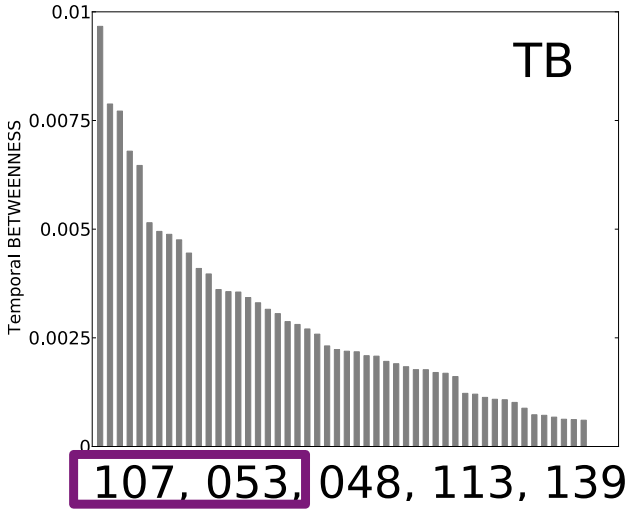
- Core 151 users
- Window size= 1 business day
- 1137 days

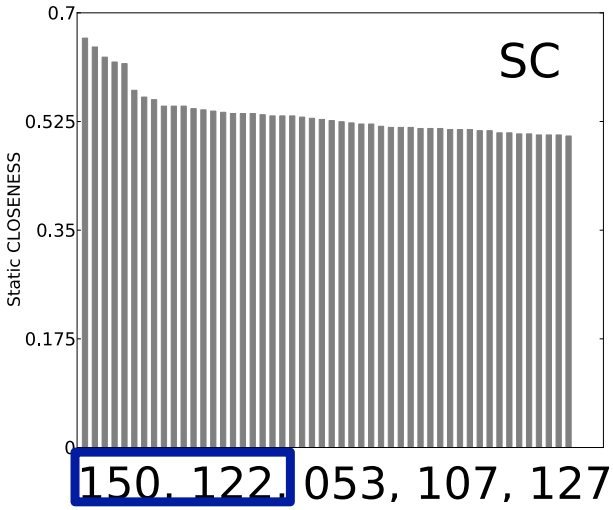


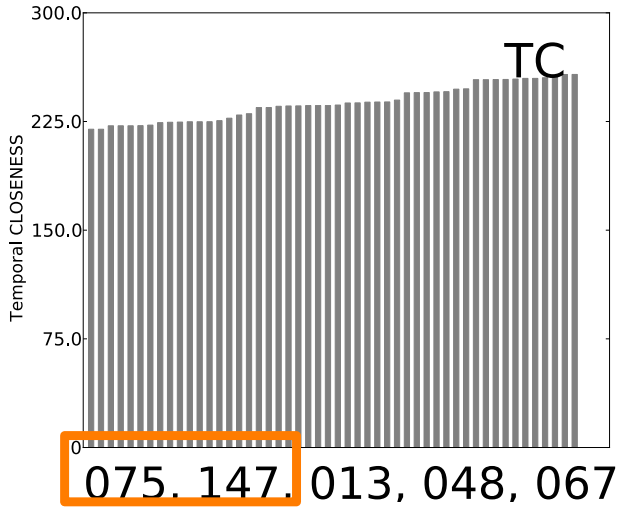


Semantics





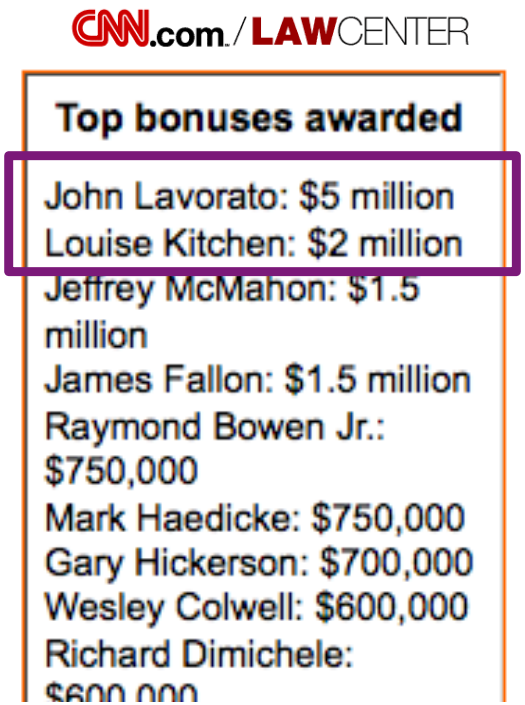




ID	Role
009	(Unknown)
013	Legal
017	Manager
048	Executive
053	Trader
054	President
067	Vice President
073	Trader
075	Director of Trading
107	Trader
122	Managing Director
127	Manager
139	Director
147	Trader
150	Secretary

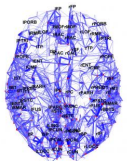
Semantics

ID	Name	Role
9	Stephanie Panus	(Unknown)
13	Marie Heard	Legal
17	Mike Grigsby	Manager
48	Tana Jones	Executive
53	John Lavorato	Trader
54	Greg Whalley	President
67	Sara Shackleton	Vice President
73	Jeff Dasovich	Trader
75	Gerald Nemec	Director of Trading
107	Louise Kitchen	Trader
122	Sally Beck	Managing Director
127	Kenneth Lay	Manager
139	Mary Hain	Director
147	Carol Clair	Trader
150	Liz Taylor	Secretary



- Big bonuses linked with information mediators

Small-world Behaviour in Time-Varying Networks



Brain network

Bluetooth (INFOCOM'06)

Facebook London Network

	C	C^{rand}	L	L^{rand}	E	E^{rand}
α	0.44	0.18	3.9 (100%)	4.2 (98%)	0.50	0.48
β	0.40	0.17	6.0 (94%)	3.6 (92%)	0.41	0.45
γ	0.48	0.13	12.2 (86%)	8.7 (89%)	0.39	0.37
δ	0.44	0.17	2.2 (100%)	2.4 (92%)	0.57	0.56
d1	0.80	0.44	8.84 (61%)	6.00 (65%)	0.192	0.209
d2	0.78	0.35	5.04 (87%)	4.01 (88%)	0.293	0.298
d3	0.81	0.38	9.06 (57%)	6.76 (59%)	0.134	0.141
d4	0.83	0.39	21.42 (15%)	15.55(22%)	0.019	0.028
Mar	0.044	0.007	456	451	0.000183	0.000210
Jun	0.046	0.006	380	361	0.000047	0.000057
Sep	0.046	0.006	414	415	0.000058	0.000074
Dec	0.049	0.006	403	395	0.000047	0.000059

What's Next?

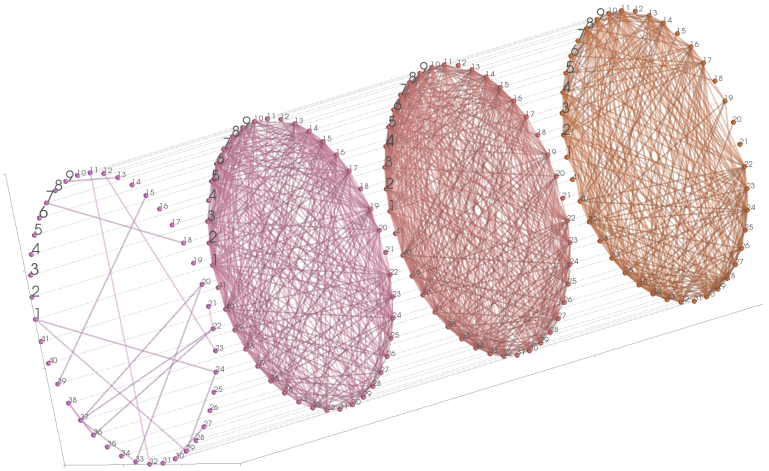
- Analysis vs Prediction
- Multi-dimensional networks
- Integration of the geographic aspects: spatio-temporal analysis
- Application to security problems
- System issues: design of scalable systems for real-time data processing

Questions?

Mirco Musolesi

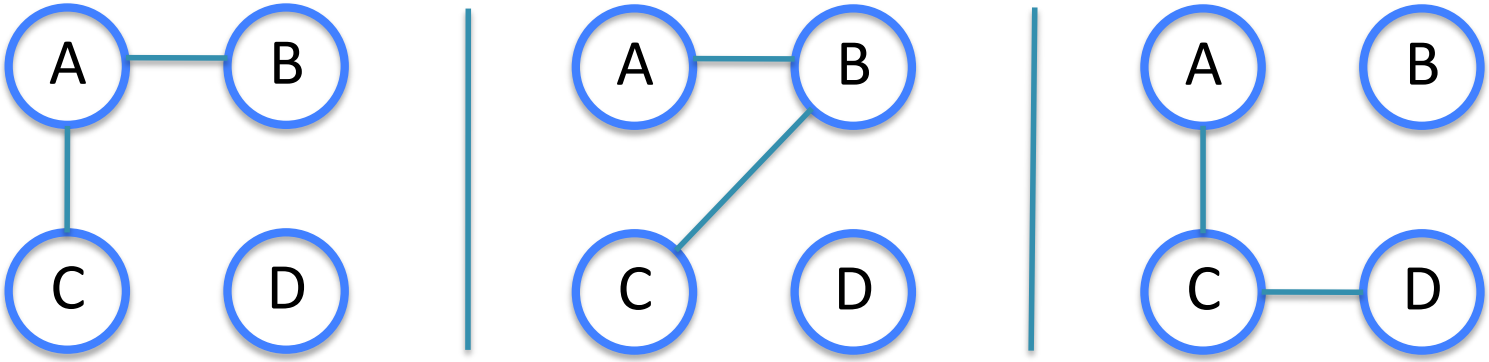
School of Computer Science

University of Birmingham

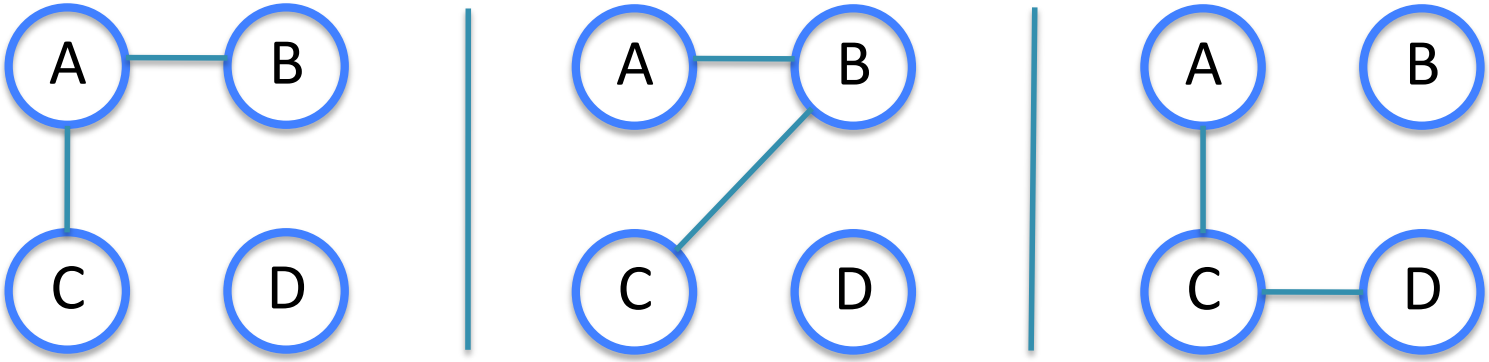


Email: m.musolesi@cs.bham.ac.uk

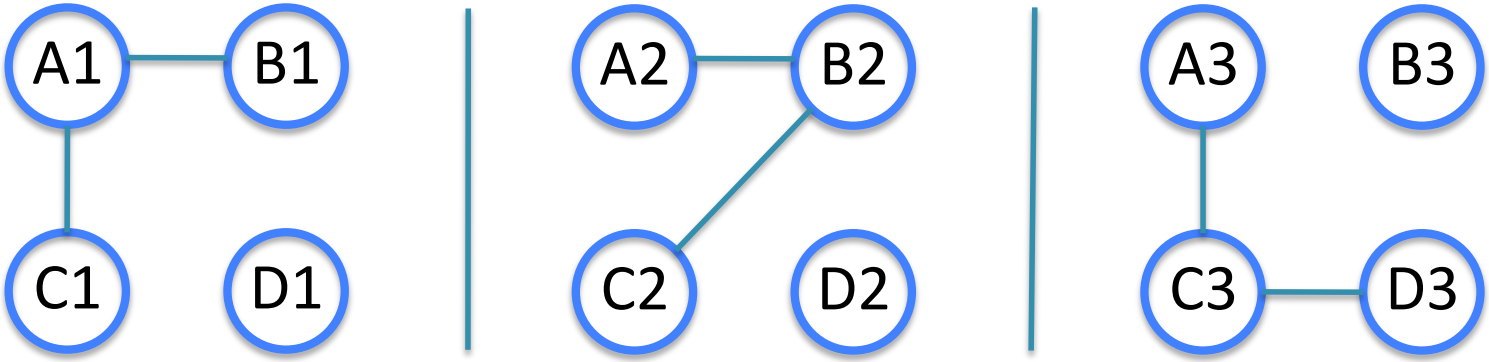
Web: http://www.cs.bham.ac.uk/~musolesm



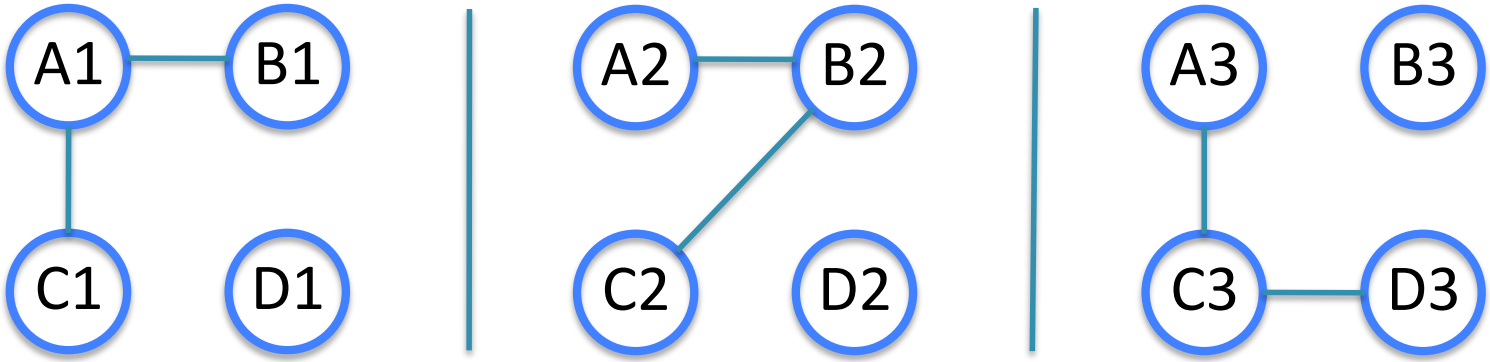
t_1 t_2 t_3



t_1 t_2 t_3



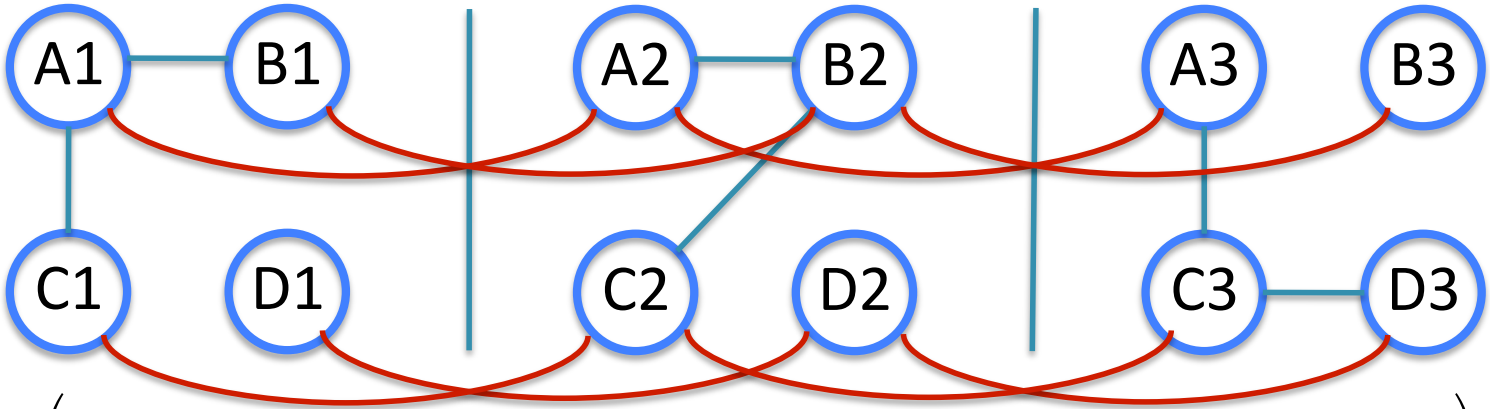
t_1 t_2 t_3



A =

	A1	B1	C1	D1	A2	B2	C2	D2	A3	B3	C3	D3
A1	0	1	1	0	1	0	0	0				
B1	1	0	0	0	0	1	0	0				
C1	1	0	0	0	0	0	1	0				
D1	0	0	0	0	0	0	0	1				
A2					0	1	0	0	1	0	0	0
B2					1	0	1	0	0	1	0	0
C2					0	1	0	0	0	0	1	0
D2					0	0	0	0	0	0	0	1
A3									0	0	1	0
B3									0	0	0	0
C3									1	0	0	1
D3									0	0	1	0

Mirco Musolesi



A =

	A1	B1	C1	D1	A2	B2	C2	D2	A3	B3	C3	D3
A1	0	1	1	0	1	0	0	0				
B1	1	0	0	0	0	1	0	0				
C1	1	0	0	0	0	0	1	0				
D1	0	0	0	0	0	0	0	1				
A2					0	1	0	0	1	0	0	0
B2					1	0	1	0	0	1	0	0
C2					0	1	0	0	0	0	1	0
D2					0	0	0	0	0	0	0	1
A3									0	0	1	0
B3									0	0	0	0
C3									1	0	0	1
D3									0	0	1	0

Temporal Katz Centrality

$$\boldsymbol{W} = \boldsymbol{I} + \alpha \boldsymbol{A} + \alpha^2 \boldsymbol{A}^2 + \ldots = \sum_{n=0}^{\infty} \alpha^n \boldsymbol{A}^n$$

w_{ij}

α i j

\+

α^2 i j

\+

α^3 i j

\+

...

Temporal Katz Centrality

$$\boldsymbol{W} = \sum_{n=0}^{\infty} \alpha^n \boldsymbol{A}^n = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1}$$

$$\alpha < \frac{1}{\lambda_1}$$

Largest eigenvalue

$$c_i = \sum_j w_{ij}$$

$$\boldsymbol{c} = \boldsymbol{W} \cdot \boldsymbol{1}$$

Katz Centrality

Computational Issues

$$\boldsymbol{c} = \boldsymbol{W} \cdot \boldsymbol{1}$$

$$\boldsymbol{c} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1} \cdot \boldsymbol{1}$$

Complexity related to the calculation of the inverse matrix is $O((T*N)^3)$ where N is the number of nodes and T is the number of the time windows.

$$\boldsymbol{c} = \alpha \boldsymbol{A}\boldsymbol{c} + \boldsymbol{1}$$

This form can be used to calculate an approximate solution by iteration. We first use an initial estimate of c and then repeat the process until convergence. The complexity of this algorithm is $O((T*N)*\#\ Iterations)$