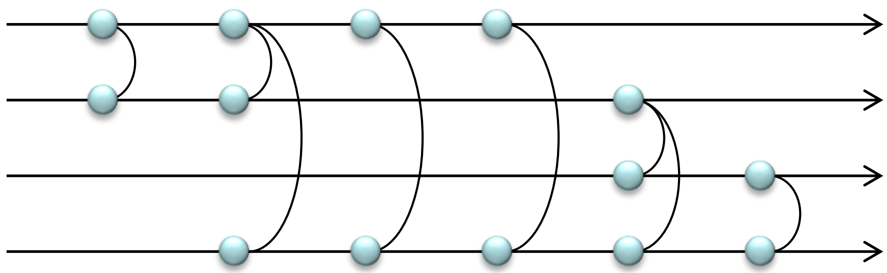
Temporal Networks



Hiroki Sayama
sayama@binghamton.edu

Temporal networks

- Networks whose topologies and activities change over time
- Heavily data-driven research
  - E.g. human contacts (email, social media, physical proximity)
- Tools still under active development
  - We will need to code a lot ourselves!

Real-world temporal networks

- Social (face-to-face) contacts
- Email transactions among employees
- Data transactions in distributed computer systems
- Dynamically changing food webs

Example: SocioPatterns.org

SocioPatterns

follow us on twitter

ABOUT | GALLERY | PUBLICATIONS | NEWS | PRESS | DATA

WELCOME

SocioPatterns is an interdisciplinary research collaboration formed in 2008 that adopts a data-driven methodology to study social dynamics and human activity.
To this aim, we have developed infrastructures based on small wireless sensors that interact over short distances and can detect the physical proximity and face-to-face contacts of the individuals wearing them. We have used it to gather longitudinal data on human mobility and face-to-face proximity in real-world environments, covering widely varying contexts across several countries: schools, museums, hospitals, etc. We use the data to study human behaviour and to develop agent-based models for the transmission of infectious diseases.

We make most of the collected data freely available to the scientific community.

FEATURED: INFECTIOUS SOCIOPATTERNS POSTER

NEWS

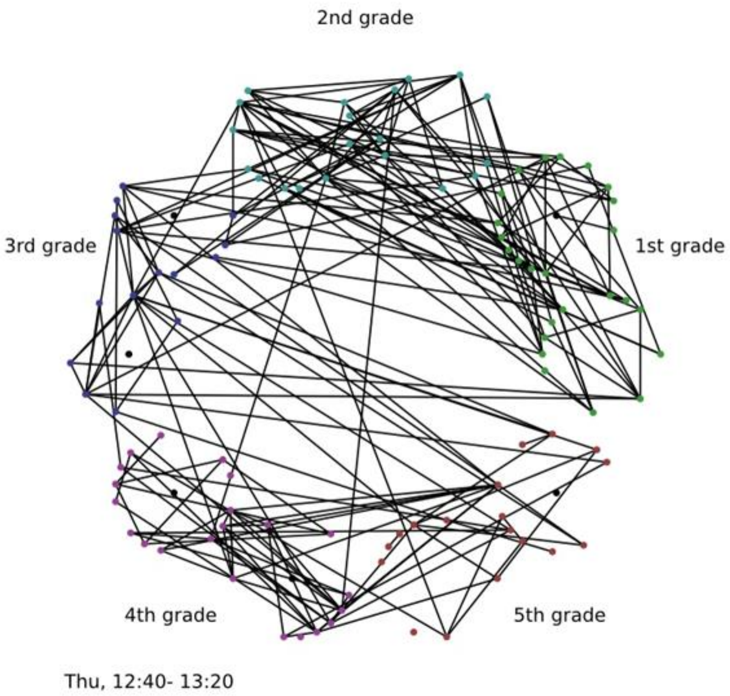
New paper in Nature Communications

We have published a new paper in Nature Communications. In this paper, we consider the issue of how to correctly inform numerical models of the propagation of infectious diseases when only partial information on the contacts of individuals is available, due to population sampling.
Indeed, the coverage of the population in many measures of detailed contact networks is incomplete, and this yields a systematic underestimation of epidemic risk if the data is used without precaution. Here, we introduce a method to compensate for this systematic bias and obtain accurate evaluations of epidemic risk using incomplete data. To this aim, we have developed an algorithm that effectively fills in the gaps of the empirical data with a realistic picture of the missing contact network. Although the obtained surrogate contacts are different from the actual missing contacts, using them in the simulation of an influenza-like process gives an accurate estimation of what would have been obtained on using complete data. It is therefore possible to have a good estimation of the epidemic risk, even if a substantial fraction of the contacts are missing from the empirical data.

Example

- Dynamic contact patterns in a primary school

  https://vimeo.com/31490438



Representations of Temporal Networks

General representation

- To represent a temporal network, define the following function:

$$a(i, j, t) = 0 \text{ or } 1 \text{ (or weight)}$$

  - $a(i, j, t) = 1$ if node i is connected to node j at time t, otherwise 0

  (If time t is discrete, this could be represented by an adjacency tensor)

Two popular classes of temporal networks with continuous time

- Temporal networks with contact sequence
  - Edges appear momentarily at certain time points and then disappears
- Temporal networks with connection intervals
  - Edges exist for certain time periods

Contact sequence

i	j	t
(1,	2,	1)
(1,	2,	2)
(1,	4,	2)
(1,	4,	3)
(1,	4,	4)
(2,	3,	5)
(2,	4,	5)
(3,	4,	6)

…

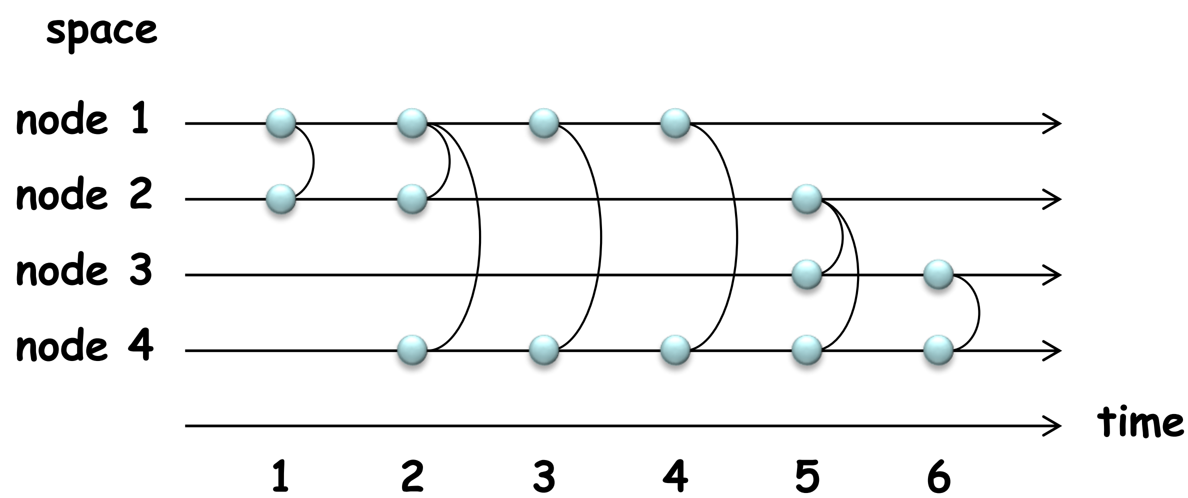
Edge list with contact times

{
(1, 2): (1, 2, …),
(1, 4): (2, 3, 4, …),
(2, 3): (5, …),
(2, 4): (5, …),
(3, 4): (6, …)
}

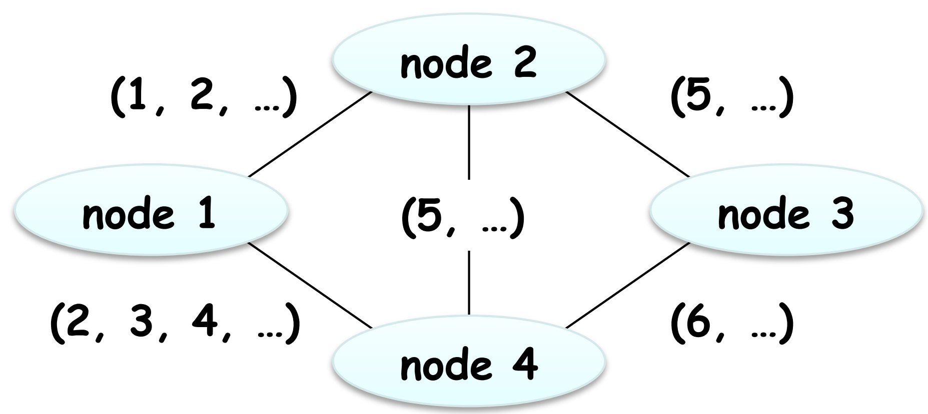
(Time does not have to be discrete)

Contact sequence: Visualized

(1, 2, 1)
(1, 2, 2)
(1, 4, 2)
(1, 4, 3)
(1, 4, 4)
(2, 3, 5)
(2, 4, 5)
(3, 4, 6)
…



{
(1, 2): (1, 2, …),
(1, 4): (2, 3, 4, …),
(2, 3): (5, …),
(2, 4): (5, …),
(3, 4): (6, …)
}



Connection intervals

i	j	t_1	t_2
(1,	2,	0.5,	1.0)
(1,	2,	1.5,	2.5)
(1,	4,	2.0,	4.0)
(2,	3,	3.5,	5.0)
(2,	4,	5.0,	6.0)
(3,	4,	5.5,	6.5)

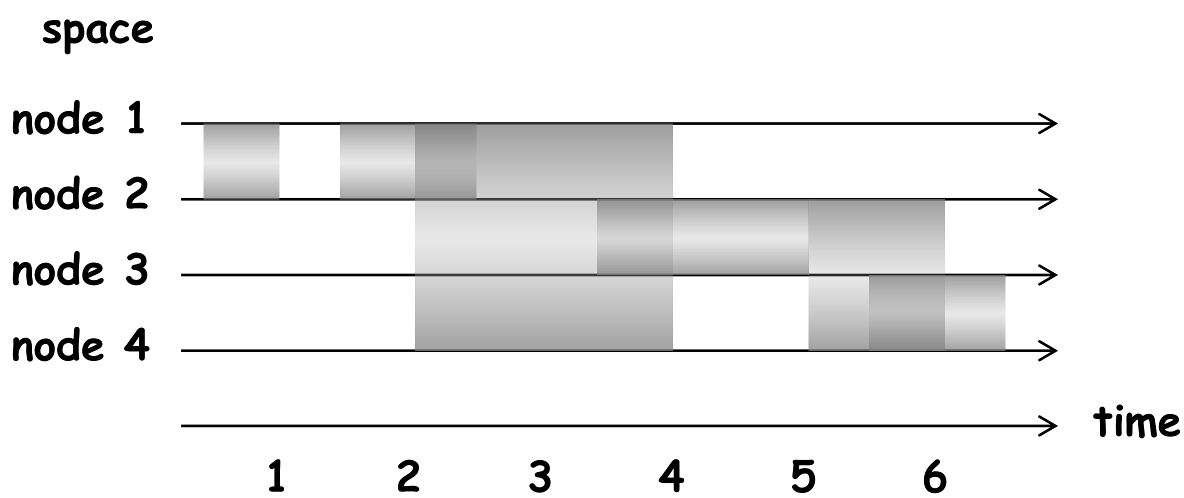
…

Edge list with connection intervals

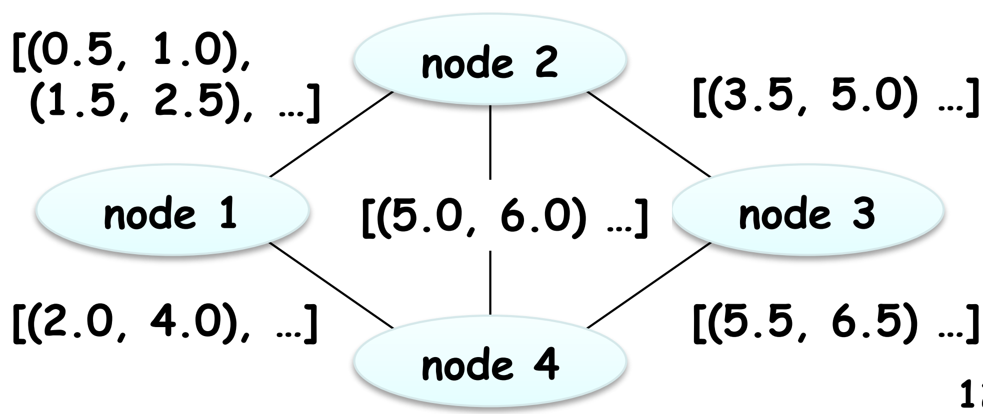
```
{
  (1, 2): [(0.5, 1.0),
           (1.5, 2.5), …],
  (1, 4): [(2.0, 4.0), …],
  (2, 3): [(3.5, 5.0) …],
  (2, 4): [(5.0, 6.0) …],
  (3, 4): [(5.5, 6.5) …]
}
```

Connection intervals: Visualized

(1, 2, 0.5, 1.0)
(1, 2, 1.5, 2.5)
(1, 4, 2.0, 4.0)
(2, 3, 3.5, 5.0)
(2, 4, 5.0, 6.0)
(3, 4, 5.5, 6.5)
…



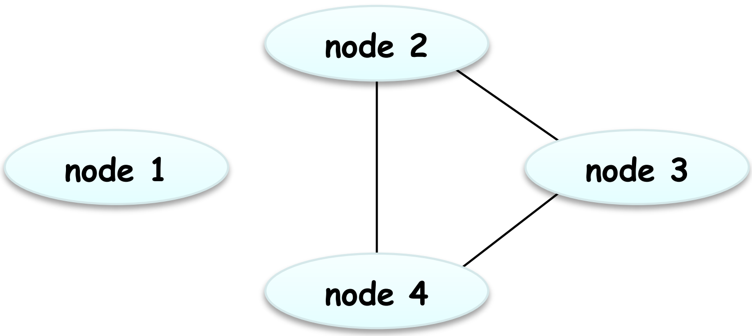

```
{
    (1, 2): [(0.5, 1.0),
             (1.5, 2.5), …],
    (1, 4): [(2.0, 4.0), …],
    (2, 3): [(3.5, 5.0) …],
    (2, 4): [(5.0, 6.0) …],
    (3, 4): [(5.5, 6.5) …]
}
```



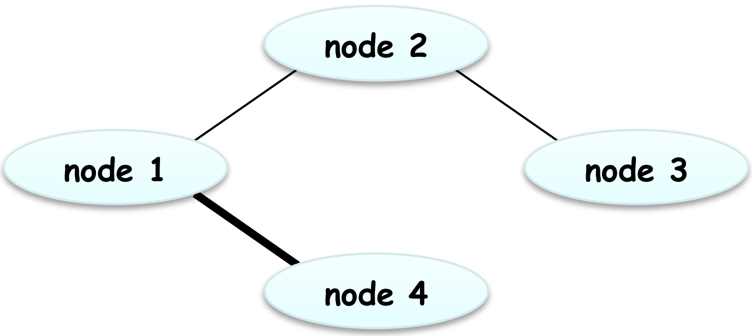
Time-aggregated network

- These networks can also be created for a specific time window (e.g., each hour, day, week, etc.)

  → Time-aggregated network



t = 5, 6



t in [2.0, 4.0]

Seeking real-world data

SocioPatterns

follow us on twitter

ABOUT | GALLERY | PUBLICATIONS | NEWS | PRESS | DATA

WELCOME

SocioPatterns is an interdisciplinary research collaboration formed in 2008 that adopts a data-driven methodology to study social dynamics and human activity.
To this aim, we have developed infrastructures based on small wireless sensors that interact over short distances and can detect the physical proximity and face-to-face contacts of the individuals wearing them. We have used it to gather longitudinal data on human mobility and face-to-face proximity in real-world environments, covering widely varying contexts across several countries: schools, museums, hospitals, etc. We use the data to study human behaviour and to develop agent-based models for the transmission of infectious diseases.

We make most of the collected data freely available to the scientific community.

FEATURED: INFECTIOUS SOCIOPATTERNS POSTER

NEWS

New paper in Nature Communications

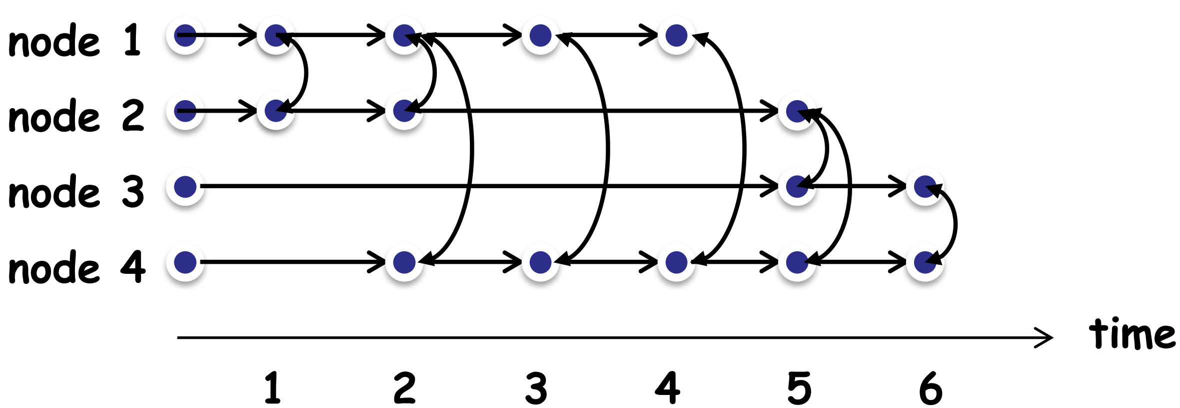
We have published a new paper in Nature Communications. In this paper, we consider the issue of how to correctly inform numerical models of the propagation of infectious diseases when only partial information on the contacts of individuals is available, due to population sampling.
Indeed, the coverage of the population in many measures of detailed contact networks is incomplete, and this yields a systematic underestimation of epidemic risk if the data is used without precaution. Here, we introduce a method to compensate for this systematic bias and obtain accurate evaluations of epidemic risk using incomplete data. To this aim, we have developed an algorithm that effectively fills in the gaps of the empirical data with a realistic picture of the missing contact network. Although the obtained surrogate contacts are different from the actual missing contacts, using them in the simulation of an influenza-like process gives an accurate estimation of what would have been obtained on using complete data. It is therefore possible to have a good estimation of the epidemic risk, even if a substantial fraction of the contacts are missing from the empirical data.

Exercise

- Download the "Primary school temporal network data" and look at its contents
- Read its contact sequence data into Python
- Generate and visualize an aggregated network on a specific hour

Plotting contacts in space-time

1. Create space-time nodes at t = 0 for all nodes
   - This may be skipped if you prefer not to have them
2. Read a new contact
   - If it occurs later than the time points of involved nodes, create new space-time nodes for them, add directed edges from old ones, and update their time points
3. Create edges between involved nodes
4. Go back to 2



Exercise

- Plot the contacts in the primary school temporal network data in detail over space-time (just for the hour chosen in the previous exercise)
  - How to lay out space-time nodes?

Exercise: Dynamic animation

- You could also dynamically visualize the contacts using PyCX's pycxsimulator.py

- Visualize the behavior of the primary school temporal network data using PyCX

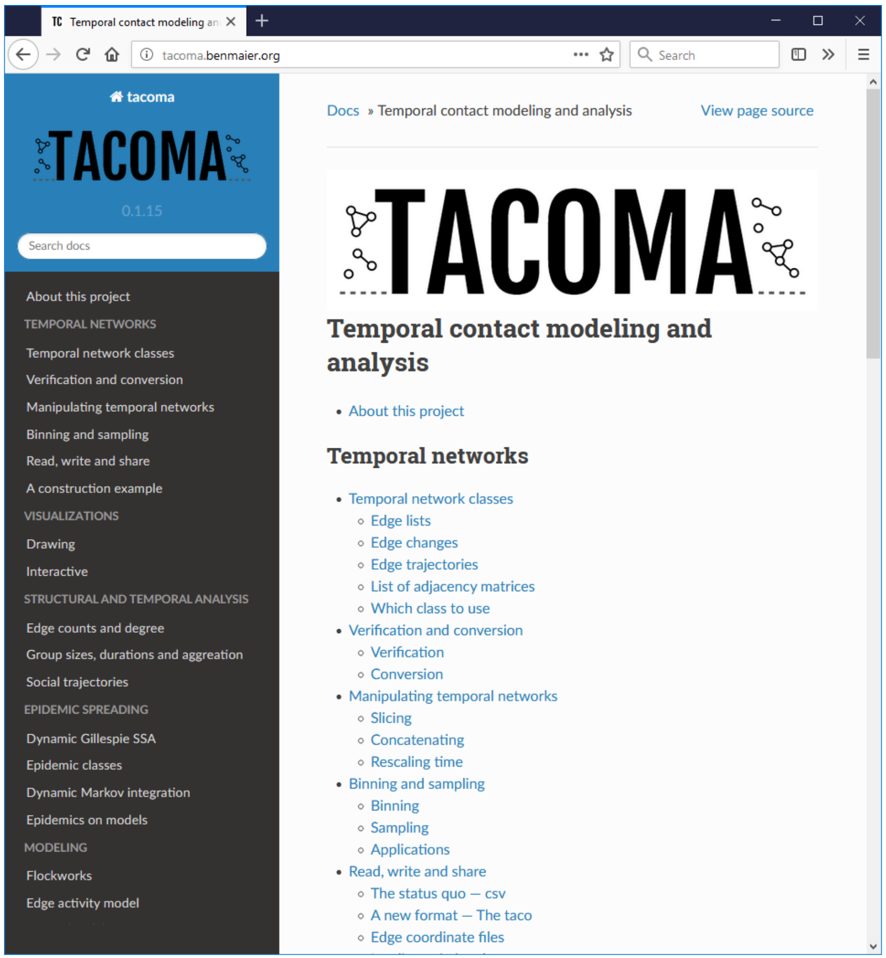
Exercise

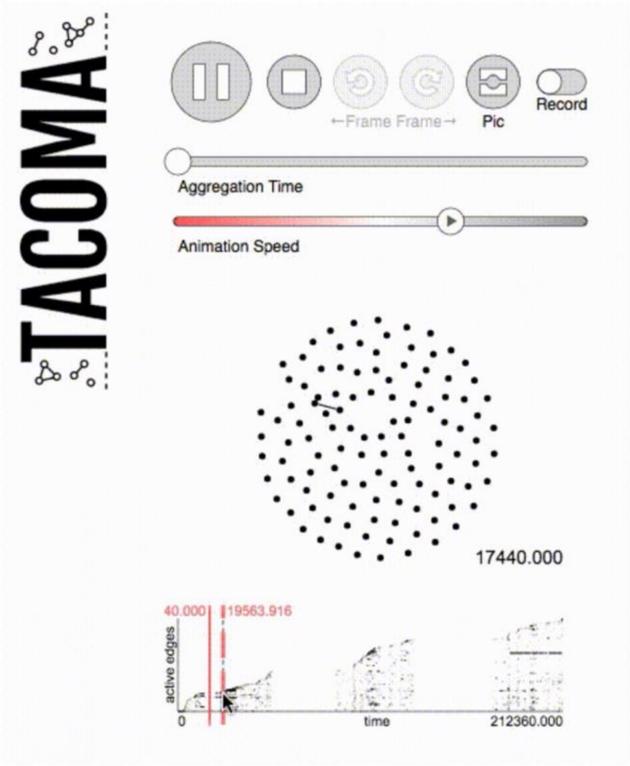
- Calculate the number of contacts in the primary school temporal network data at each time point and plot it over time

- Plot the distribution of time intervals between contacts on edge (1437, 1563) (= most active edge)

- What do you see there?

New tool: TACOMA



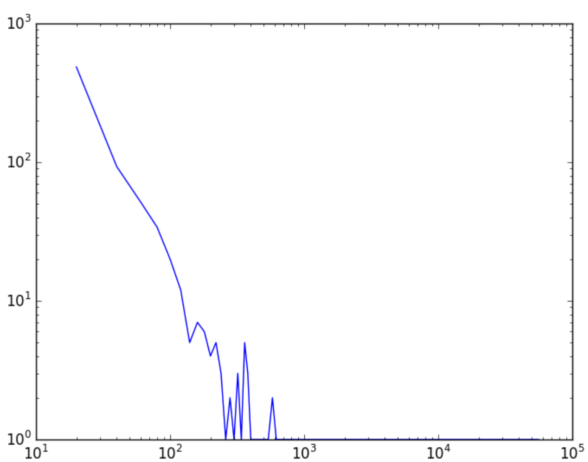


http://tacoma.benmaier.org/

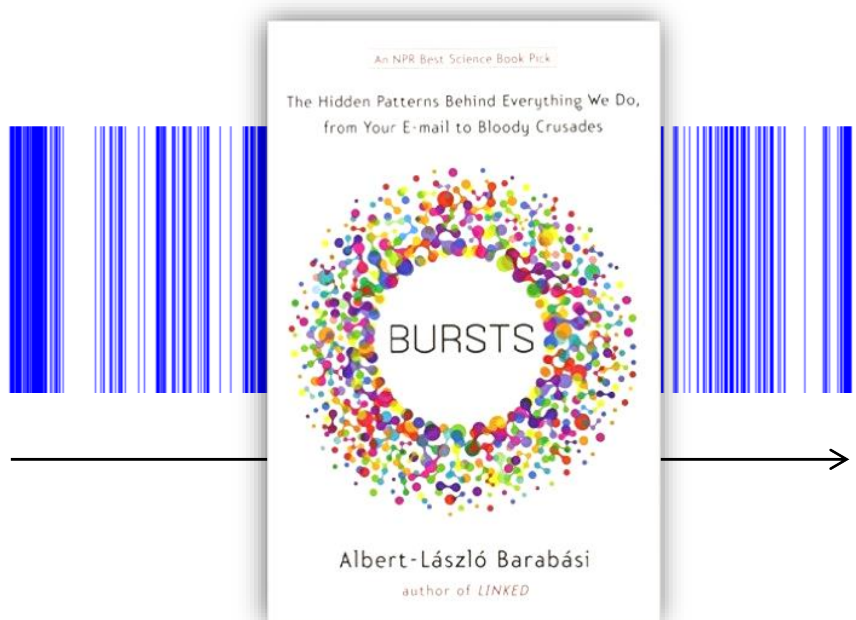
Burst!

- Many real-world social contacts show "bursty" behavior
  - Intervals between two consecutive contacts show a power-law distribution

frequency



interval length



time

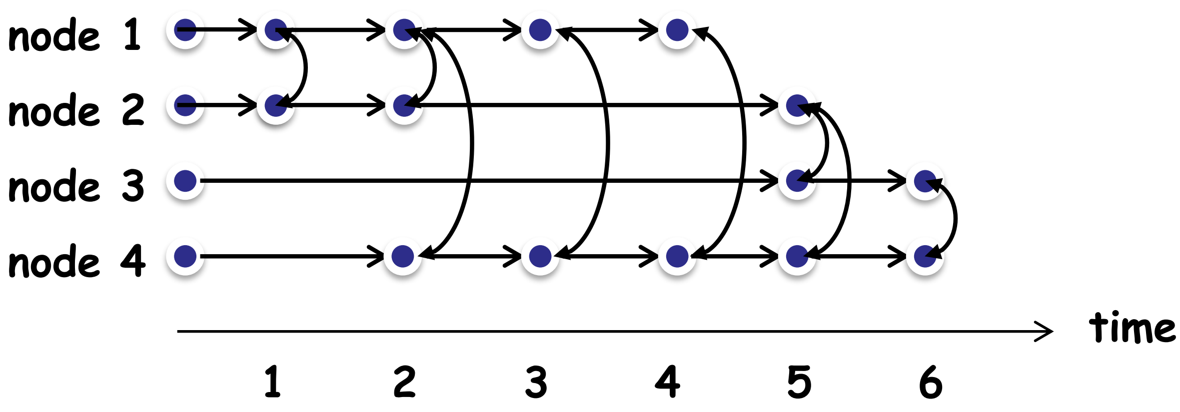
Implications of bursts

- Events are interrelated over time too
- Temporal behavior of the system may not be fully captured using time averages or Poisson distributions
- Dynamics on the network (e.g., disease spreading) may be fundamentally different from what would be expected on non-bursty networks

Time-Respecting Paths

Time-respecting path

- A sequence of contacts with non-decreasing times
  - E.g. There are time-respecting paths from node 1 to 3 but not the other way



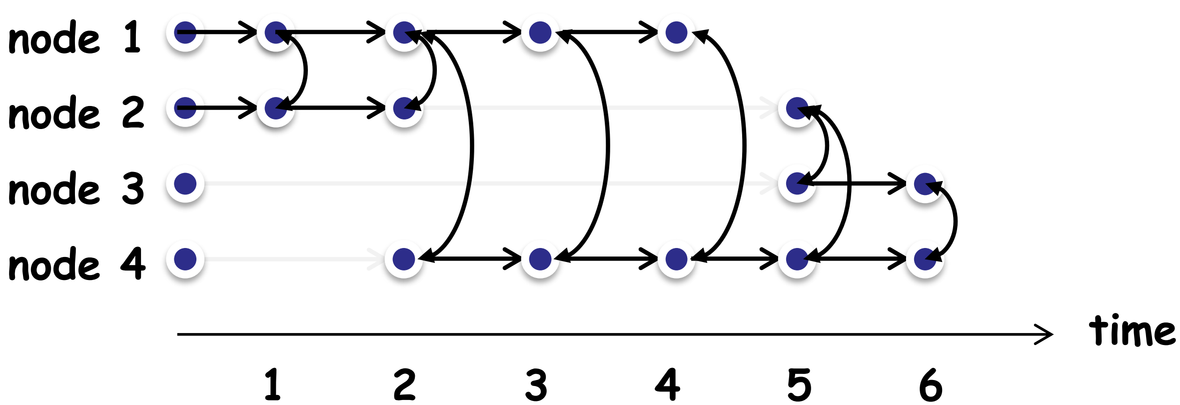
- It can be obtained simply as a path in the space-time representation

Exercise

- Check if there are any time-respecting paths between randomly chosen two nodes in the space-time network constructed previously

Time-respecting path with limited waiting time

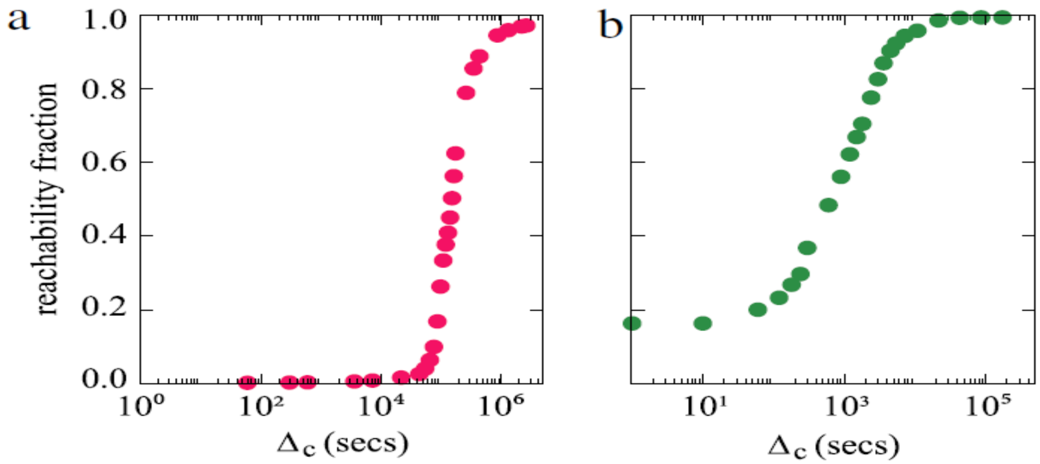
- Time-respecting path that does not contain same-node edges longer than a certain threshold (waiting time)
  - Represents loss of infectivity, etc.



(Not allowed to stay on the same node for >1 time step)

Reachability

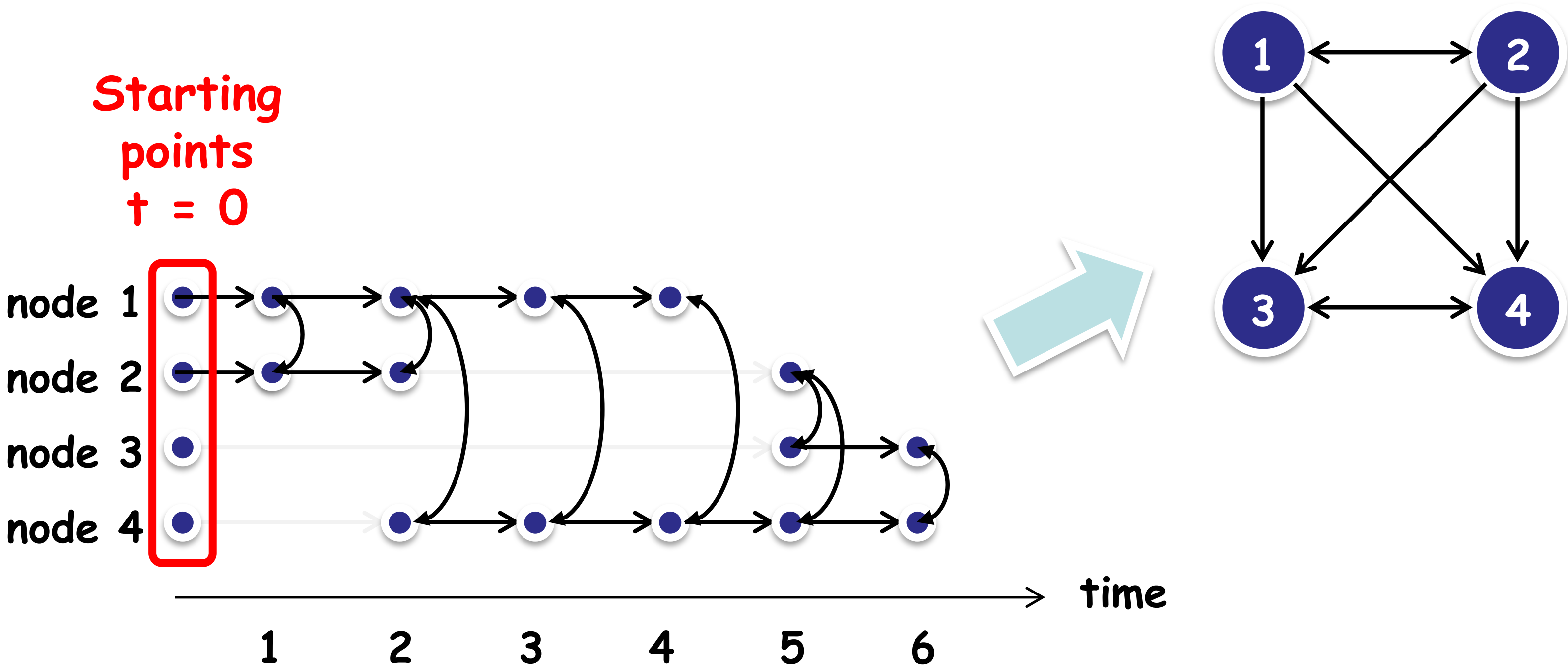
- Where can you reach from a given node at a specific time point, following only time-respecting paths?
  - Within a certain time window
  - With limited waiting time



Reachability ratios (a: mobile phone call network, b: airline connection network)
Source: Holme & Saramäki (2012)

Reachability network

- A static network in which reachable nodes are connected by edges



Exercise

- Draw the reachability network of the primary school temporal network data during the hour chosen in the previous exercises
  - With max. waiting time = 1 minute
  - What happens if you increase the limit?

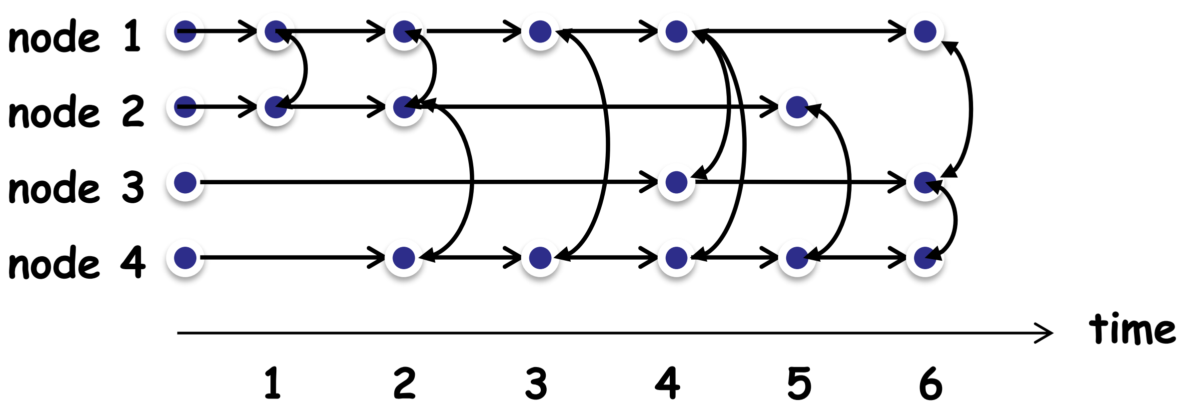
Temporal Network Measurements

Latency

- Length of time needed to move from node i to node j using the *fastest* time-respecting path (within a given time window)
  - Use the "fastest", not necessarily the "shortest" path
- Average latency captures how quickly spreading is occurring in the network

Exercise

- What is the latency between each pair of nodes? Average?



- What if the time window is [4, 6]?

Exercise

- Calculate the average latency of the primary school temporal network data during each hour
- Plot how the average latency changes over time

Centralities

- Measurable on aggregated networks
- Can be generalized to temporal networks by replacing:
  - Degree by node activity
  - Shortest path by fastest time-respecting path (for closeness, betweenness)
- Simulating propagation of importance through temporal edges gives centralities similar to eigenvector ones

Exercise

- Calculate the temporal closeness centralities of nodes in the primary school temporal network data during a certain hour
  - This can be calculated by averaging latencies from the focal node to all other nodes

Exercise

- Simulate propagation of importance on the primary school temporal network data with the following assumptions:
  - Each node initially has importance 1
  - Each time two nodes have a contact, 10% of their respective importance is moved to the other side through the contact
  - This repeats for the entire data set

Randomization of Temporal Networks

Randomization of temporal networks

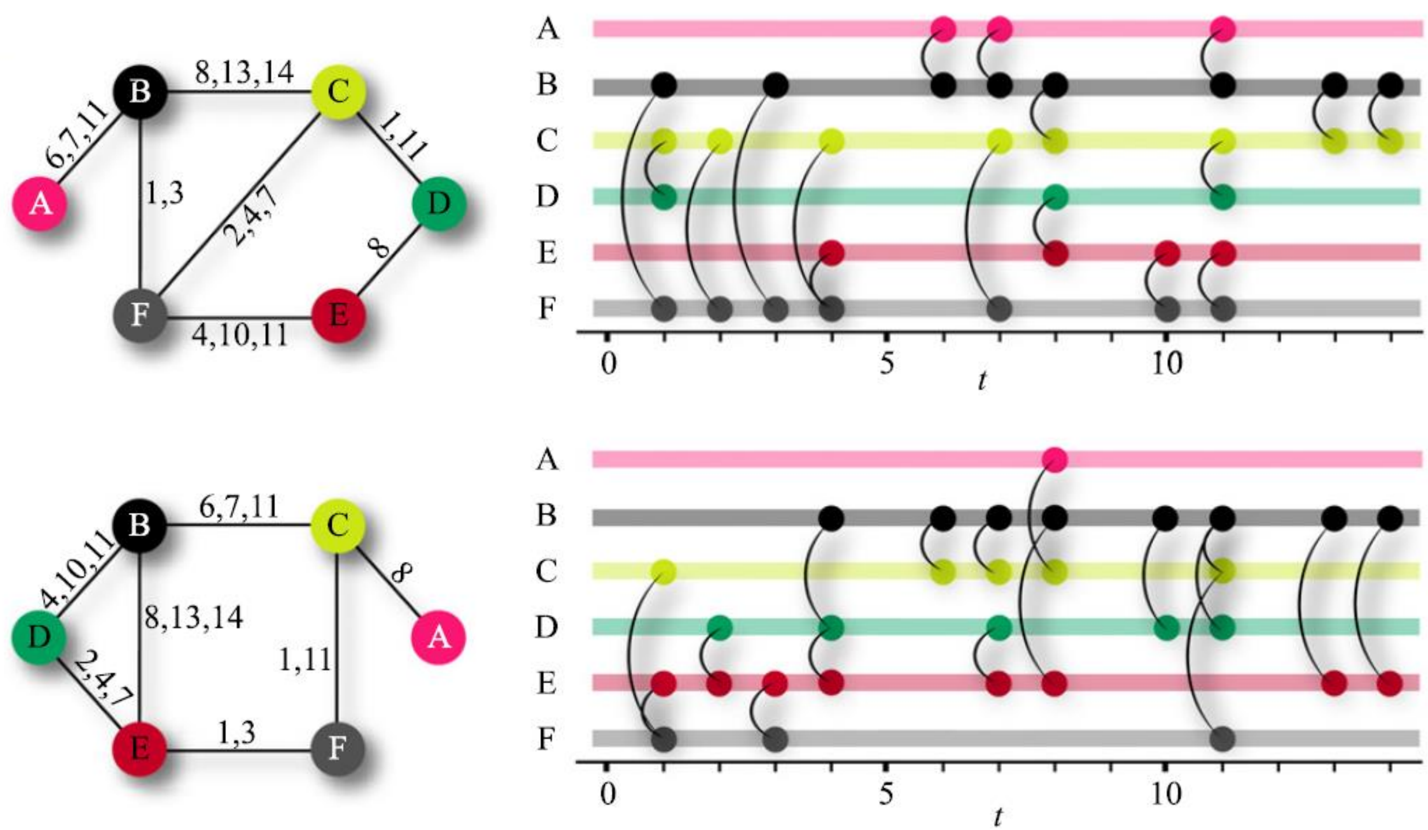
- Two main approaches:

  - Spatial randomization
    - E.g. randomized edges

  - Temporal randomization
    - E.g. randomly permuted times

Randomized edges (RE)

- Randomizes the topology of the aggregated network by double edge swaps, etc.



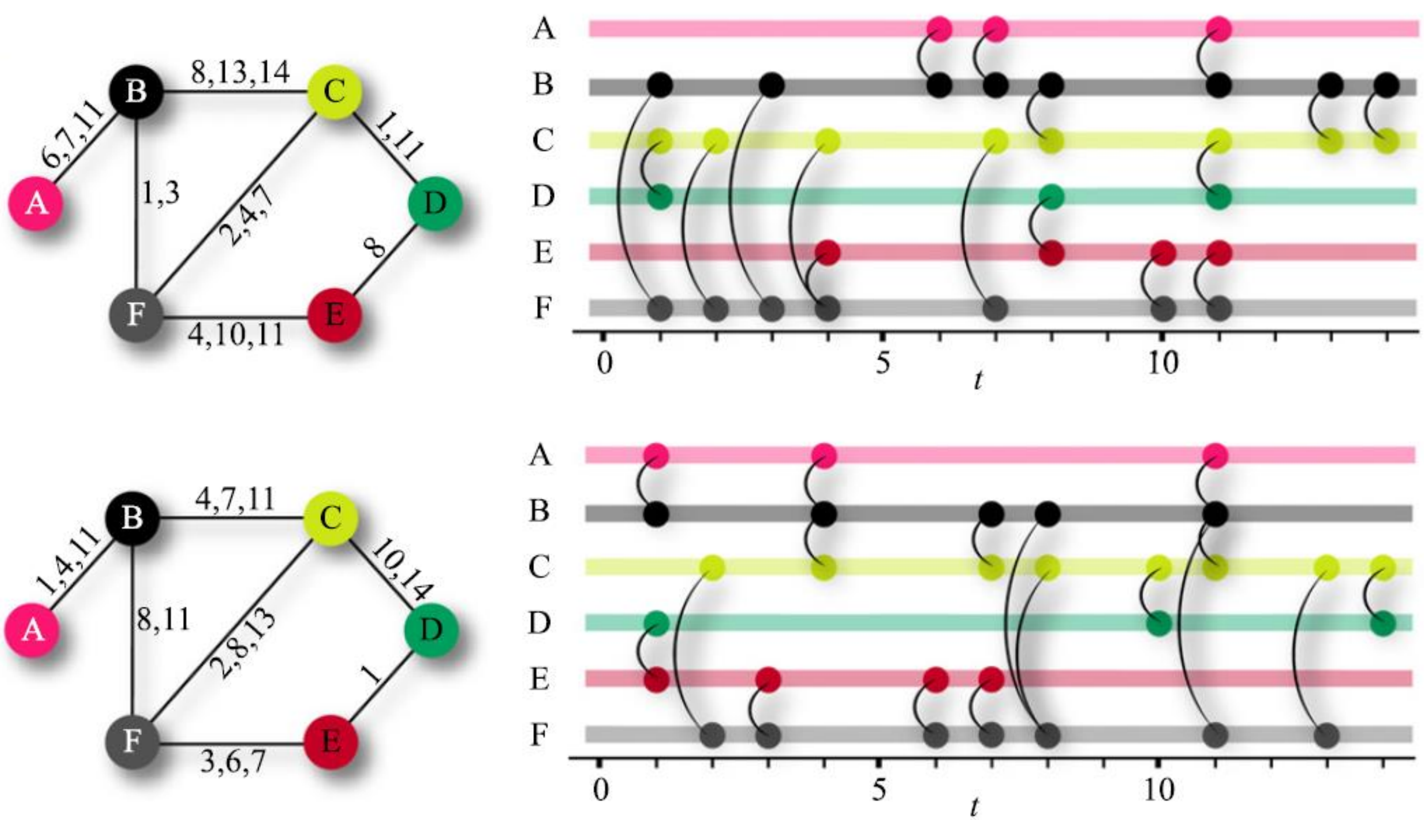
Source: Holme & Saramäki (2012)

Properties of RE method

- RE destroys spatial (topological) correlations
- Preserved properties
  - Node degrees (but not # of contacts)
  - Contact sequences
  - Activity dynamics (both micro and macro)
- Works for both contact sequences and connection intervals

Randomly permutated times (RP)

- Randomly reshuffles time stamps of contacts



Source: Holme & Saramäki (2012)

Properties of RP method

- RP destroys temporal correlations
- Preserved properties
  - Network topology
  - Number of contacts on each node/edge
  - Macroscopic activity dynamics (but not burstiness on individual edges)
- Very easy to implement
  - Needs overlap check for networks with connection intervals

Other randomization methods

- RE + RP
  - Randomizes both space and time
  - Preserves only macro activity dynamics
- Random times (RT)
  - Time stamps of contacts are sampled randomly from a uniform distribution
  - Destroys temporal correlations and also macroscopic activity dynamics
- RE + RT etc…

Exercise

- Implement several randomization methods (RE, RP, RE+RP, RT, RE+RT, etc.)
- Apply them to the primary school temporal network data
- Measure some network properties and see how each randomization affects
- Interpret the results

Dynamical Processes on Temporal Networks

Dynamics on temporal networks

- Temporal networks are particularly suitable for studying the dynamics of spreading processes
  - Infectious diseases via social contacts
  - Computer viruses via email transactions
  - Information spreading on social media

- How do the properties of the network affect the spreading process?

Example: SIS model

- Infection rate per contact: p_i
- Recovery rate per unit of time: p_r

- Epidemic process can be simulated on empirically observed temporal network data
  - Dynamics of contact formation could also be simulated using a stochastic model (not discussed much in this class)

Exercise

- Simulate the SIS model on the primary school temporal network data
- Obtain parameter values (p_i, p_r) with which epidemics occurs

- Then, randomize the network data using RE, RP, RE+RP, RT, RE+RT, etc.
  - How does this affect simulation results?