Time-varying networks approach to face-to-face empirical interactions

Michele Starnini

Departament de Fisica i Enginyeria Nuclear, UPC, Barcelona, Spain

Barcelona,
06/02/2014

Thesis advisor: Romualdo Pastor-Satorras
Work done in collaboration with ...

• Andrea Baronchelli, City University London (UK)
• Alain Barrat, CNRS Aix Marseille Université (France)
• Ciro Cattuto, ISI Foundation (Italy)
• Anna Machens, CNRS Aix Marseille Université (France)
• Romualdo Pastor Satorras, Universitat Politècnica de Catalunya (Spain)
Outline

• Motivation: Time-varying Networks and Human activity patterns

• Face-to-face empirical interactions: Data gathering and Modeling

• Diffusive Processes on Temporal Networks: Random Walks and Epidemic Spreading
compute their coordinates using a variation of the mapping method that requires only local topological information (see Supplementary Methods). In Figure 7a, we show the performance of greedy forwarding in the resulting maps at each time step, and observe only minor performance degradation, even over long time scales. In a nutshell, the existing AS coordinates are essentially static, as once computed they can stay the same for years.

Existing Internet topology measurements including the Archipelago data are known to be incomplete and miss some AS links. Therefore, a natural question is how this missing information affects the quality of the constructed map, and the performance of greedy forwarding in it. Intuitively, as the performance of greedy forwarding is robust with respect to link removals, we might expect it to be robust with respect to missing links as well. Moreover, if the constructed map is used in practice, then greedy forwarding will see and use those links that topology measurements do not see. We might thus also intuitively expect greedy forwarding to perform better in practice than we report in this section, simply because those missing links, when used by greedy forwarding, would provide additional shortcuts between potentially remote ASs. We confirm this intuition in Figure 7b with experiments emulating the missing link issue.

The success ratio degrades only slowly as a function of the fraction of missing links, whereas if we add the emulated missing links back, then the success ratio increases as expected. Therefore, the routing results reported here should actually be considered as lower bounds for greedy routing performance that can be achieved in practice using the constructed hyperbolic Internet map.

Discussion

We have constructed a hyperbolic map of the Internet, and release this map as part of the Supplementary Data set. The map can be used for essentially infinitely scalable Internet routing. The amount of routing information that ASs must maintain is proportional to the AS degree, which is theoretically best possible as ASs must always keep some information about their neighbours. Routing communication overheads are also minimized, as ASs do not exchange any routing information on dynamic changes of the AS topology.

The presented solution thus achieves routing efficiency that is theoretically...
...or among us: social networks

Actors co-starring movies

"Eyes Wide Shut" (1999)

Bela Lugosi
Nicole Kidman
Tom Cruise

"Plan 9 from Outer Space" (1959)

Mona McKinnon
...or among us: social networks

A. Actors co-starring movies
   - Tom Cruise
   - Nicole Kidman
   - "Eyes Wide Shut" (1999)

B. Scientists co-writing papers
   - Bela Lugosi
   - Mona McKinnon
   - "Plan 9 from Outer Space" (1959)

(not really the same...)
Patterns of human activity

Which is the distribution $P(\tau)$ of time gaps $\tau$ between two consecutive events?

Classical models based in Poisson process are contradicted by empirical observations:

$$P(\tau) \sim \tau^{-\gamma}$$

Human activities follow a bursty behavior.
In the context of human dynamics, the distribution of time intervals between consecutive tasks, such as e-mails or printer usage, often follows a power law. This has been observed in both human-initiated activities, like the correspondence patterns of Einstein and Darwin, and computer-driven environments, such as web browsing and printer usage.

For human-initiated activities, the distribution of time intervals between e-mails and printer usage is typically characterized by a power law with an exponent close to 1. This suggests a bursty nature in these activities, indicating that while the means have changed, the distribution remains similar to that observed in the historical record of letter writing.

In computer-driven environments, such as web browsing, the distribution of file sizes and response times also follows a power law, suggesting that the nature of human dynamics is a consequence of a queuing process where users need to decide what to do next, and the execution of tasks is not random but biased towards higher-priority items.

The critical regime of human dynamics is particularly relevant for understanding the nature of human behavior, as it is a consequence of a queuing process with nonrandom task selection. This provides insights into the fundamental and potentially generic features of human behavior.
Real networks are dynamic entities, links are rewired on various time scales.

The temporal dimension impacts the dynamical processes developing on networks.

Social networks are intrinsically dynamic, interactions begin and end constantly.
The reproductive number identifies the average number of secondary cases generated by a typical infected individual. In an homogeneous population, the behavior of the epidemics is controlled by the reproductive number. In the past few years, the inclusion of complex networks and mobility schemes into the substrate of spreading processes such as contagion, diffusion, transfer, etc. has highlighted new and interesting results on the propagation mechanisms and dynamical processes in social and biological systems.

The epidemic threshold is inversely proportional to the second moment of the network's degree distribution. In particular, for networks characterized by a fixed, heterogeneous mean-field approaches are coupled together. Let us assume a distribution of activity potential according to a specific time sequence defined by the agents' activity. By working with activity rates, we can derive epidemic evolution out relying on any time-aggregated view of the network connectivity. Several results highlight the importance of considering all edges as always available to carry the contagion process, disregarding the fact that the edges may be active or not.

The above finding can be more precisely quantified by calculating the epidemic threshold in activity-driven networks with a different average degree of the networks. In particular, for networks characterized by a fixed or annealed per capita spreading rate $\beta$, empirically measured by using four different time windows and a schematic representation of the model. Considering just 13 nodes and a susceptible of class $a$, the third term on the right side takes into account the probability that the epidemic threshold depends on the topological properties of the networks. In particular, in panel (A) we show the cumulative distributions of the observables $G$ and $R$.

In Fig. 4-B we plot the results of numerical simulations of the SIS model on a network generated according to our model and aggregated over all time steps. We observe that the two aggregated instances of the SIS model on a network generated according to our model and of the activity potential, $G$, empirically measured by using four different time windows and a schematic representation of the model. Considering just 13 nodes and a susceptible of class $a$, the third term on the right side takes into account the probability that the epidemic threshold depends on the topological properties of the networks. In particular, in panel (A) we show the cumulative distributions of the observables $G$ and $R$.

In Fig. 4-B we plot the results of numerical simulations of the SIS model on a network generated according to our model and aggregated over all time steps. We observe that the two aggregated instances of the SIS model on a network generated according to our model and of the activity potential, $G$, empirically measured by using four different time windows and a schematic representation of the model. Considering just 13 nodes and a susceptible of class $a$, the third term on the right side takes into account the probability that the epidemic threshold depends on the topological properties of the networks. In particular, in panel (A) we show the cumulative distributions of the observables $G$ and $R$.

In Fig. 4-B we plot the results of numerical simulations of the SIS model on a network generated according to our model and aggregated over all time steps. We observe that the two aggregated instances of the SIS model on a network generated according to our model and of the activity potential, $G$, empirically measured by using four different time windows and a schematic representation of the model. Considering just 13 nodes and a susceptible of class $a$, the third term on the right side takes into account the probability that the epidemic threshold depends on the topological properties of the networks. In particular, in panel (A) we show the cumulative distributions of the observables $G$ and $R$.
SocioPatterns
Face-to-face interaction networks
Data Gathering
Empirical Data of Social Dynamics

Records of face-to-face interactions

Social Contexts:

25c3
25th Chaos Communication Congress

eswc
European Semantic Web Conference

hosp
Lyon Hospital

ht
Hypertext Conference

school
High School

sfhh
Société Française d’Hygiène Hospitalière Congress
SocioPatterns

1:00 - 2:30 PM

lunch, coffee and start of 3rd session
From face-to-face interactions to dynamical networks

Empirical data with fine-grained spatial (\(\sim 1-2\text{m}\)) and temporal (\(\sim 20\text{ sec.}\)) resolution
Human Dynamics Properties

Duration of conversations

Inter-contacts gap times

Heterogeneity of the interactions

(1 time step = 20 seconds)

Human interactions are bursty!
Aggregate Social Network

Duration of conversations

Inter-contacts gap times

Integrating Information

Heterogeneity persists over longer time intervals

Total contact time between pairs

Strength vs Degree

High connected individuals spend more time in each interaction
My first conference

Maybe that guy is more attractive than me???
A Model of Social Interactions

- N agents performs a biased random walk in a 2D space
- Whenever 2 agents intercept within a distance $d$, they start to interact

A Model of Social Interactions

• Agents can be in an active (move and interact) or inactive (not moving neither interacting) state

• From time to time, agents jump from active to inactive state with probability $r_i \in [0, 1]$ and vice versa

$\text{Inactive} \quad \xrightarrow{r_i} \quad \text{Active}$

$\quad \xleftarrow{1 - r_i}$

A Model of Social Interactions

- Each agent $i$ is characterized by his attractiveness $a_i \in [0, 1]$
- At each time step $t$ each $i$ agent moves with prob. $p_i(t) = 1 - \max_{j \in N_i(t)} a_j$

You decide if keep interacting depending on the attractiveness of your most interesting peer

A Model of Social Interactions

- N agents performs a biased random walk in a 2D space
- Agents can be in a active or inactive state
- Interactions are ruled by the attractiveness of the agents

Simple but very realistic assumptions, reproducing experimental setting

Results are robust with respect to variations of the density $\rho$. 

Statistical properties of social interactions:

- Distribution of the contact duration
- Distribution of the gap times between consecutive conversations
• The model output for the integrated weighted network is OK
• Tendency of an agent to interact with new peers decreases in time, \( k(t) \sim t^\mu, \quad \mu \approx 0.6 \)
Group dynamics

Interacting group of 5 individuals

- Probability of finding a group of size $s$
**Group dynamics**

- Probability of finding a group of size $s$

- Probability that a group of size $n$ has a duration $\Delta t$

Interacting group of 5 individuals

Big groups are less stable in data, at odds with model behavior!
Time-respecting paths

- Shortest and fastest time-respecting path length

System size of data and model are comparable
Diffusive Processes on Temporal Networks: Random Walks

How does the network time scale affect a dynamical process taking place on its top? How do the two time scales interplay?

- **Random walk (RW):** paradigm of diffusive processes, on empirical time-varying networks (SocioPatterns)
- Introduction of **randomizing strategies:** null models to check different properties of the substrate
- Importance of temporal correlation in consecutive interactions: the **burstiness of human activity**
A walker present at node $i$ at time $t$ hops to a randomly chosen neighbor.

- If the node $i$ is isolated, the walker remains at node $i$.
- In any case, time is increased $t \rightarrow t + 1$.

Dynamical networks time scale $t \rightarrow \bar{p}t$

RW is expected to move on average once every $\frac{1}{\bar{p}}$ time steps, being $\bar{p}$ probability that a node is not isolated.
**Synthetic Extensions of Data**

- **SRep** - *Sequence replication*: the sequence is repeated periodically.
- **SRand** - *Sequence randomization*: the time order of the sequence is reshuffled.
- **SSStat** - *Statistically extended sequence*: the sequence is generated by choosing randomly at each time step $\bar{n}$ conversations.
Synthetic Extensions of Data

**SRep** - Sequence replication: the sequence is repeated periodically.

**SRand** - Sequence randomization: the time order of the sequence is reshuffled.

**SStat** - Statistically extended sequence: the sequence is generated by choosing randomly at each time step $\bar{n}$ conversations.

Which properties of the original data are preserved?

- $P(\omega)$: weight distribution of the aggregated network
- $P_i(\tau)$: inter-contact gap times of a single individual $i$
- $P(\Delta t)$: contact duration distribution

<table>
<thead>
<tr>
<th>Extension</th>
<th>$P(w)$</th>
<th>$P(\Delta t)$</th>
<th>$P_i(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRep</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SRan</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>SStat</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>
Network Searching Efficiency

• **Coverage** $C(t)$: Number of different individuals visited by the walker at time $t$

• **Mean First Passage Time** (MFPT) $\tau_i$ of a node $i$:
  Average time taken by the RW to arrive for the first time at $i$

• **Reaching Probability** $P_r(i)$:
  Probability that an individual $i$ is reached by the RW in the finite sequence
Network Searching Efficiency

- **Coverage** $C(t)$: Number of different individuals visited by the walker at time $t$

- **Mean First Passage Time** (MFPT) $\tau_i$ of a node $i$:
  Average time taken by the RW to arrive for the first time at $i$

- **Reaching Probability** $P_r(i)$:
  Probability that an individual $i$ is reached by the RW in the finite sequence

For a weighted network, at mean field level (MF):

\[
\frac{C(t)}{N} \approx 1 - (1 + t/N)^{-1} \\
\tau_i = \langle s \rangle N/s_i \\
P_r(i) \approx 1 - \exp\left( -\bar{p} T s_i / N \langle s \rangle \right)
\]
Network Searching Efficiency

- **Coverage** $C(t)$: Number of different individuals visited by the walker at time $t$

- **Mean First Passage Time** (MFPT) $\tau_i$ of a node $i$:
  Average time taken by the RW to arrive for the first time at $i$

- **Reaching Probability** $P_r(i)$:
  Probability that an individual $i$ is reached by the RW in the finite sequence

For a weighted network, at mean field level (MF):

$$C(t)/N \simeq 1 - (1 + t/N)^{-1}$$
$$\tau_i = \langle s \rangle N/s_i$$
$$P_r(i) \simeq 1 - \exp(-\bar{p}Ts_i/N\langle s \rangle)$$

**MF equations apply for the SRan extension!**

The SRan extension destroys temporal correlation between consecutive contacts, thus a RW on the SRan extension is expected to behave as in the corresponding weighted projected network.
**Numerical Simulations: Coverage & MFPT**

- The **SRan** case is **well fitted** by MF prediction, **SStat** shows a **close** behavior
- The **SRep** is **considerably slower** than the others
• The SRan case is well fitted by MF prediction, SStat shows a close behavior.
• The SRep is considerably slower than the others.

<table>
<thead>
<tr>
<th>Extension</th>
<th>$P(w)$</th>
<th>$P(\Delta t)$</th>
<th>$P_i(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRep</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SRan</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>SStat</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

Temporal correlations between consecutive conversations slow down the RW exploration!
MFPT of different datasets collapse on

\[ \tau_i \sim \frac{1}{\bar{p}} \times \left( \frac{s_i}{N\langle s \rangle} \right)^{-\alpha}, \]

with exponent \( \alpha \approx 0.75 < \alpha_{MF} = 1 \)
Slow but universal behavior?

MFPT of different datasets collapse on

\[ \tau_i \sim \frac{1}{\overline{p}} \times \left( \frac{s_i}{N\langle s \rangle} \right)^{-\alpha}, \]

with exponent \( \alpha \approx 0.75 < \alpha_{MF} = 1 \)

\( P_r(i) \) of different datasets collapse but strong deviations from MF theory

\[ P_r(i)^{MF} \approx 1 - \exp(-\overline{p} T s_i / N\langle s \rangle) \]

smaller \( P_r(i) \) for nodes with high \( s_i \)
Slow but universal behavior?

MFPT of different datasets collapse on
\[ \tau_i \sim 1/\bar{p} \times (s_i/N\langle s \rangle)^{-\alpha}, \]
with exponent \( \alpha \simeq 0.75 < \alpha_{MF} = 1 \)

\( P_r(i) \) of different datasets collapse but strong deviations from MF theory
\[ P_r(i)^{MF} \simeq 1 - \exp(-\bar{p}T_s / N\langle s \rangle) \]
smaller \( P_r(i) \) for nodes with high \( s_i \)

Searching process in the empirical, correlated network is slower!
Diffusive Processes on Dynamical Networks: Epidemic spreading

How to implement a real-time vaccination for an epidemic spreading in a ongoing social event?

- A very simple SI model of epidemic diffusion, realised on empirical time-varying networks (SocioPatterns)
- Evaluation of different immunisation strategies, contrasted with benchmark protocols
- The set of individuals to immunise is determined through preliminary measurements in a training window $\Delta T$
Why this dynamical perspective is important?

Here the disease can spread from A to B and from B to C.

But contacts are not simultaneously…
Why this dynamical perspective is important?

Here the disease can spread from A to B and from B to C

But contacts are not simultaneously…

If $t_1 < t_2$ then C cannot be infected!
Measuring immunisation efficiency

\[ \tau_i = \left\langle \frac{T_j^i - T_j}{T_j} \right\rangle_{j \neq i} \]

**Infection delay ratio:** estimate the impact of immunise element \( i \) in slowing down the spreading diffusion

\( T_j^i \) and \( T_i \) are the **time needed to infect half population** for an SI process with seed \( j \), with and without immunisation of \( i \).

\( I_j^\nu \) and \( I_j \) are the **fraction of infected individuals** for an SI process with seed \( j \), with and without immunisation of \( i \).

Consider a set of individuals \( \mathcal{V} \):

**Half infection time**

\[ \tau_\mathcal{V} = \left\langle \frac{T_j^\mathcal{V} - T_j}{T_j} \right\rangle_{j \notin \mathcal{V}} \]

**Average outbreak ratio**

\[ i_\mathcal{V} = -\left\langle \frac{I_j^\mathcal{V} - I_j}{I_j} \right\rangle_{j \notin \mathcal{V}} \]
Which is the optimal immunisation strategy?

Vaccination protocols:
- Degree (K): individuals with highest aggregated degree [deterministic]
- Bet. Centr. (BC): individuals with the highest betweenness centrality [deterministic]
- Acquaintance (A): one random contact of randomly chosen individuals [stochastic]
- Weight (W): most frequent contact of randomly chosen individuals [stochastic]
- Recent (R): last contact of randomly chosen individuals [stochastic]

Benchmarks:
- Optimal (T): individuals with the highest infection delay ratio \( T_i \)
- Random (Rn): randomly chosen individuals.
How to determine the individuals to immunise?

Training window: gathering information about the individuals (degree, BC, etc…)

- First we select the individuals by exploring the training window
- Then we run SI model over the whole contact sequence
• The **simplest degree strategy** obtain the best results in slowing down the epidemic diffusion

• A **limited knowledge** of the contact sequence, $\Delta T \approx 0.2–0.3$, is sufficient to estimate which nodes have to be immunised.
What's next?

Attractiveness model of face-to-face interactions

- Testing attractiveness model against other data of human interactions (work in progress...)
- Can we find a proxy for the attractiveness? Spatial movements? Psychological profiles? Speech power?

Dynamical processes on Multiplex Networks:

- How do different layers of interactions (f2f, emails, calls, texts...) affect diffusion (i.e. gossip propagation) taking place on them?
Thank you!

