Uncovering the structure and dynamics of information flow on the Telegram network

Thomas Louf, Aurora Vindimian, Riccardo Gallotti

September 2, 2025

UNIVERSITÀ DI TRENTO
Dipartimento di Matematica

FONDAZIONE BRUNO KESSLER

Uncovering the structure and dynamics of information flow on the Telegram network

Thomas Louf, Aurora Vindimian, Riccardo Gallotti

September 2, 2025

Telegram is...

- Not just a messaging app, it hosts **millions of public channels**, some gathering tens / hundreds of thousands of users

Telegram is...

- Not just a messaging app, it hosts **millions of public channels**, some gathering tens / hundreds of thousands of users
- An online social media = large-scale, privately-owned “public” spheres → hard to limit **spread of misinformation, hate speech** and other goodies

Telegram is...

- Not just a messaging app, it hosts **millions of public channels**, some gathering tens / hundreds of thousands of users
- An online social media = large-scale, privately-owned "public" spheres → hard to limit **spread of misinformation, hate speech** and other goodies
- Increasingly **popular**, and already very popular in some countries (Russia, Iran)

so...

Telegram is...

- Not just a messaging app, it hosts **millions of public channels**, some gathering tens / hundreds of thousands of users
- An online social media = large-scale, privately-owned "public" spheres → hard to limit **spread of misinformation, hate speech** and other goodies
- Increasingly **popular**, and already very popular in some countries (Russia, Iran)

so...

$\stackrel{?}{=}$

Not exactly!

- Its **structure is based on channels**

Not exactly!

- Its **structure is based on channels**

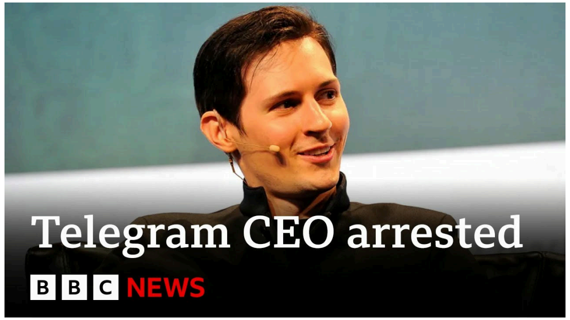
↪ information is more containerised

Not exactly!

- Its **structure is based on channels**
  - ↪ information is more containerised
  - ↪ not easy to *search* for content

Not exactly!

- Its **structure is based on channels**
  - ↪ information is more containerised
  - ↪ not easy to *search* for content
- Largely **unregulated**



Not exactly!

- Its **structure is based on channels**

  ↪ information is more containerised

  ↪ not easy to *search* for content

- Largely **unregulated**
- … and also **under-studied** scientifically

Not exactly!

- Its **structure is based on channels**
  ↪ information is more containerised
  ↪ not easy to *search* for content
- Largely **unregulated**
- … and also **under-studied** scientifically

→ Lots of work to be done

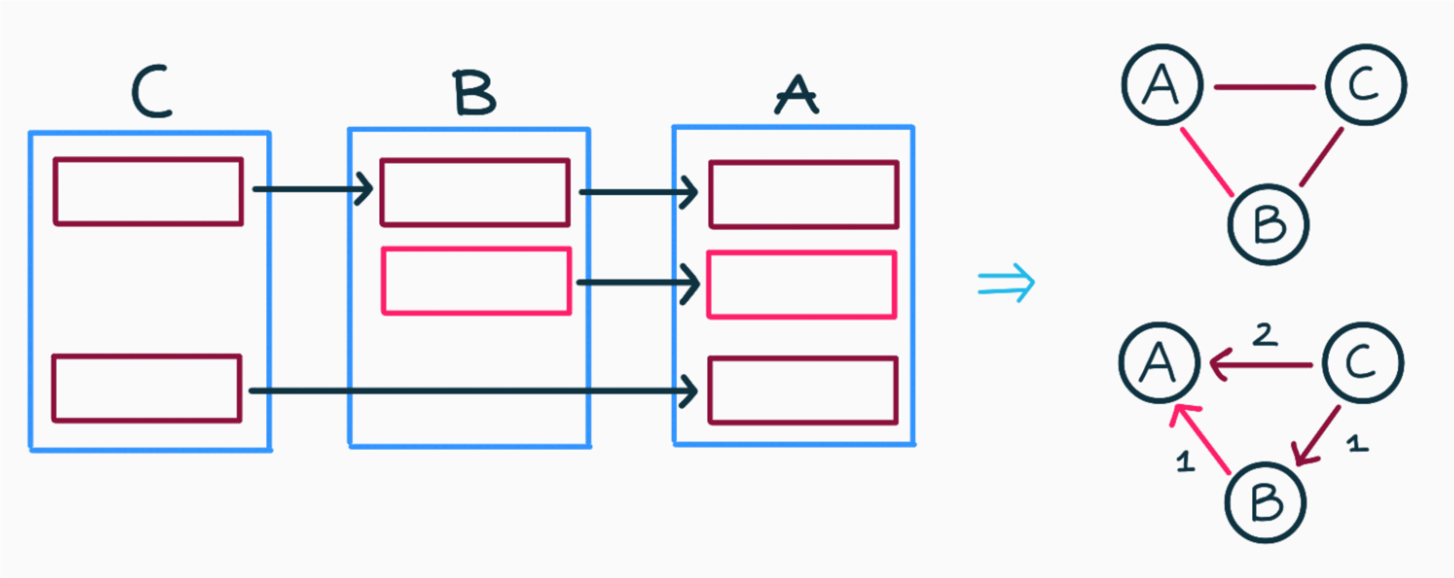
Is the network of Telegram channels anything like a social network?

What are the main mechanisms giving rise to it?

→ Analysis of the Telegram network using the Pushshift dataset (Baumgartner et al., 2020)

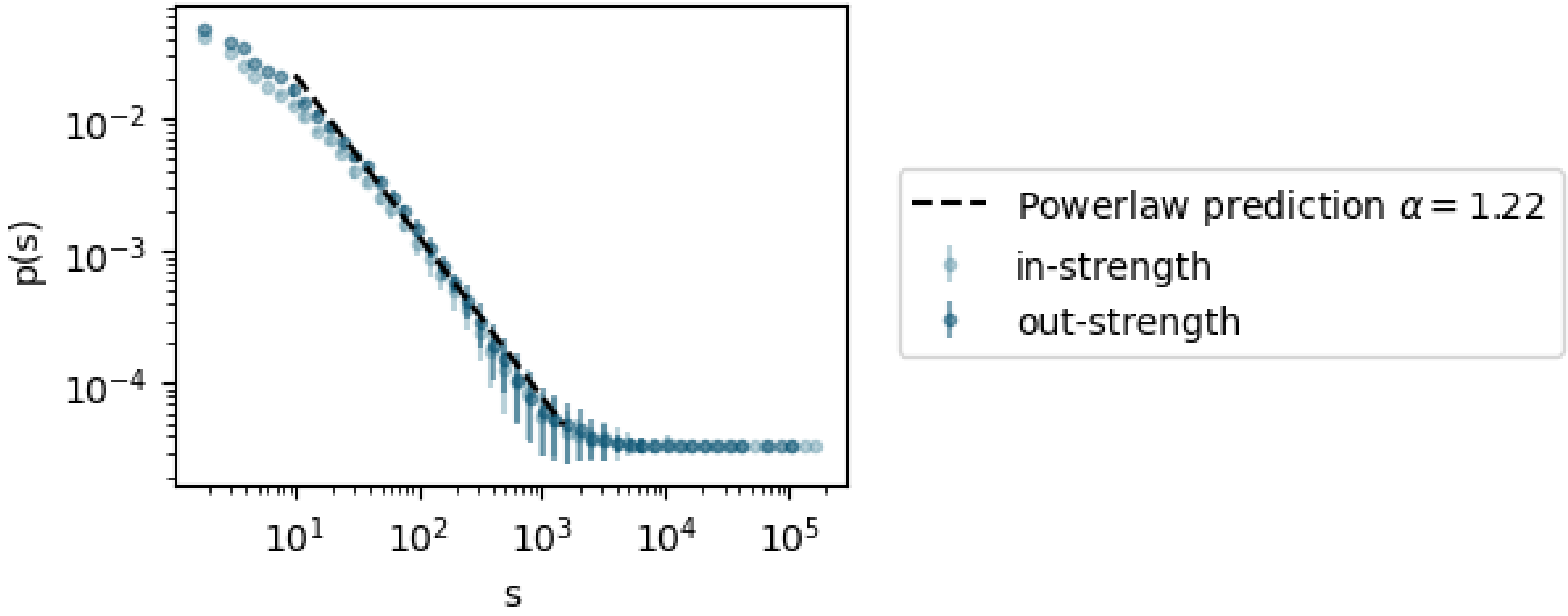
→ Model that reproduces its *topological* and *temporal* features

- Nodes: $29\,609$ channels
- Edge from B to A when A forwards a message from B → $501\,897$ directed edges

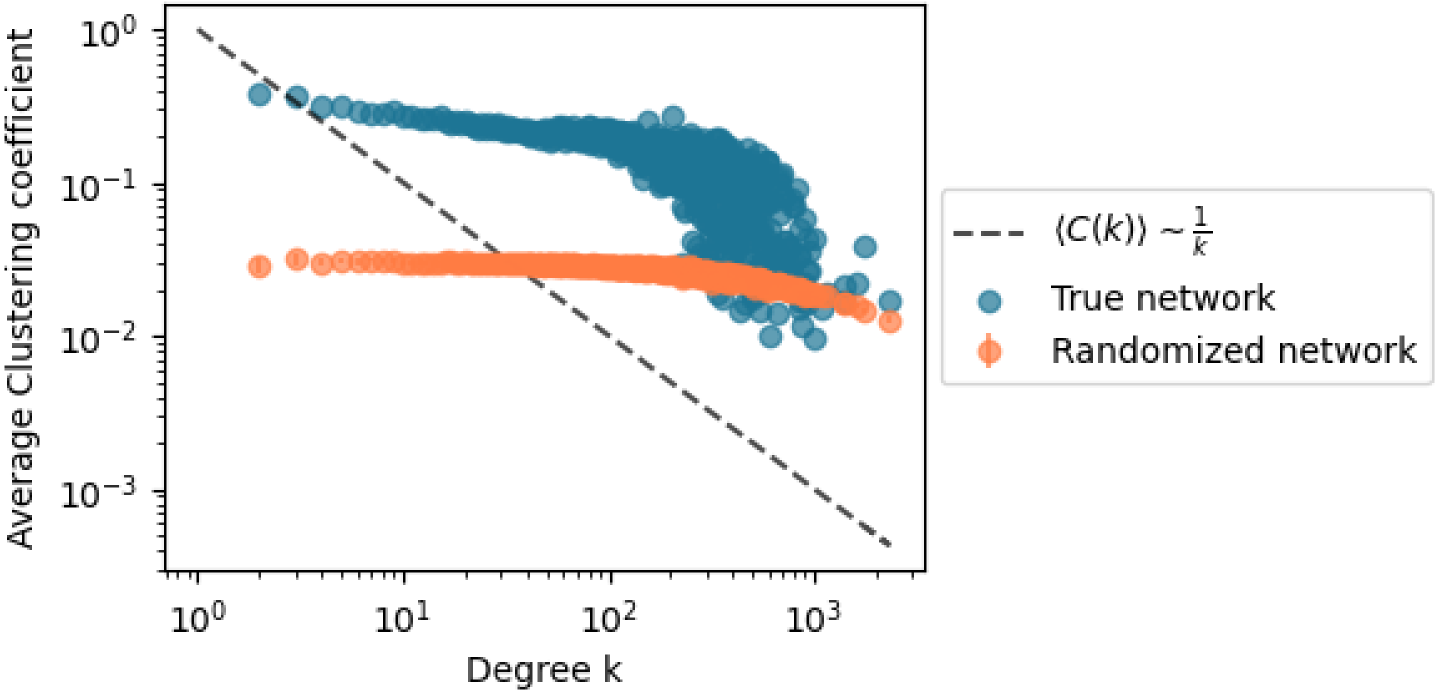


→ Network of information flow

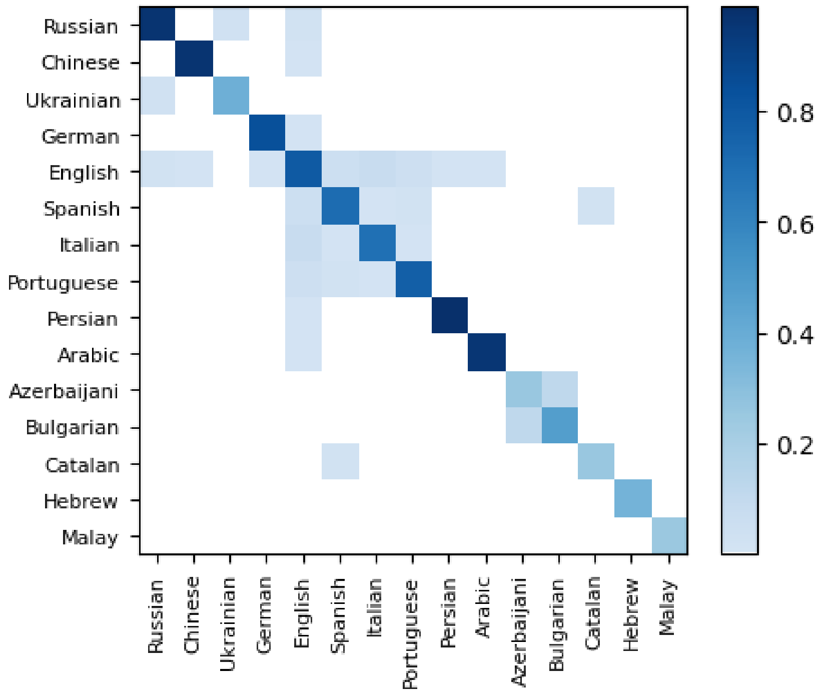
Do we have the usual Pareto law?



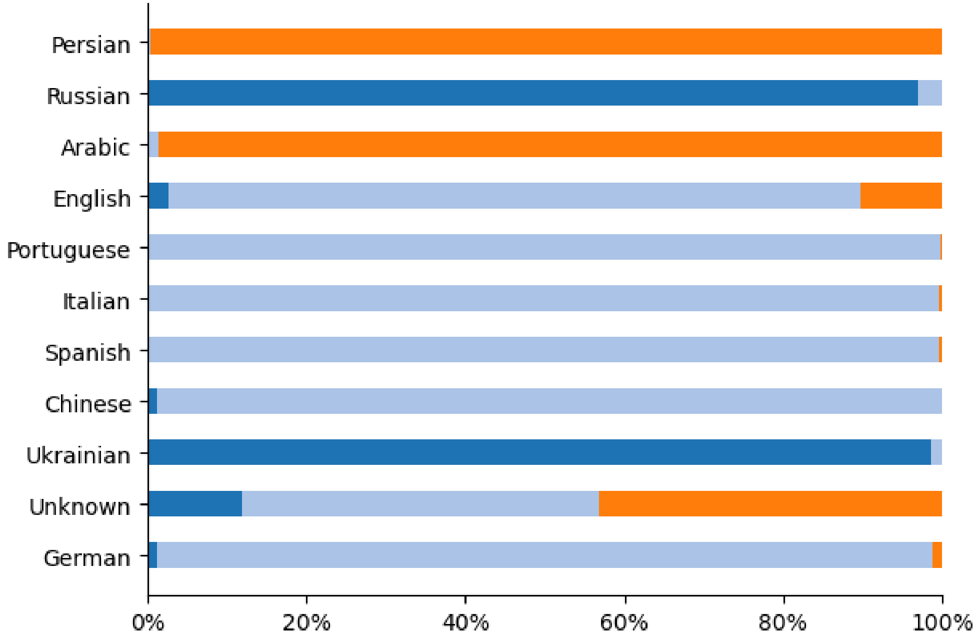
Tendency to forward from friends of my friends?



Ties formed preferably with same language...



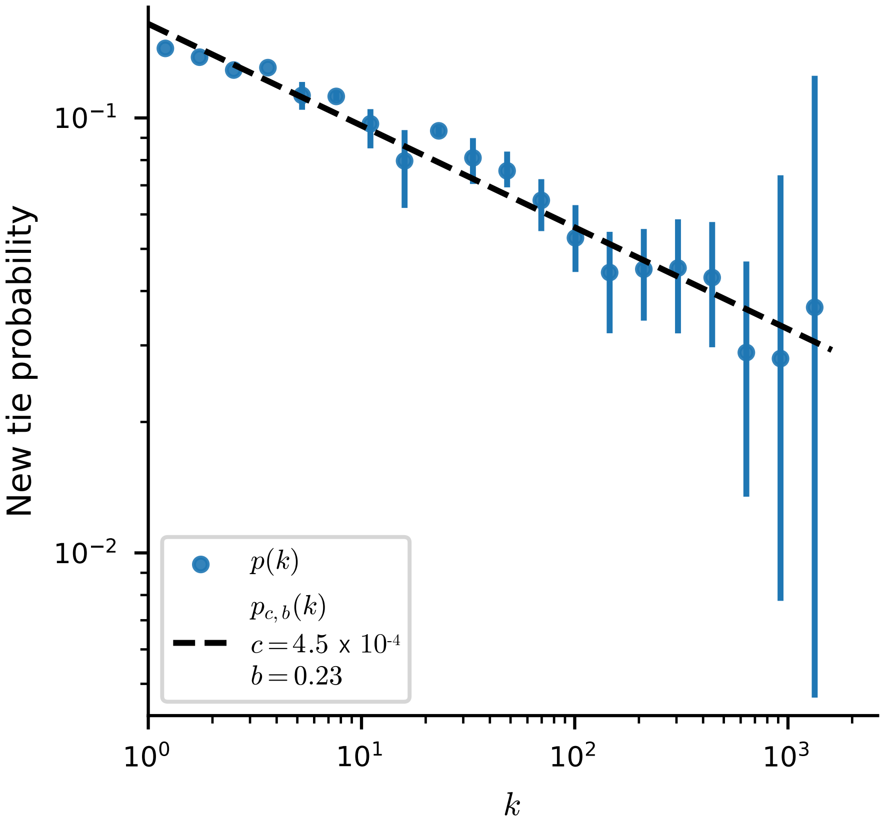
...also reflected in community partition (SBM)



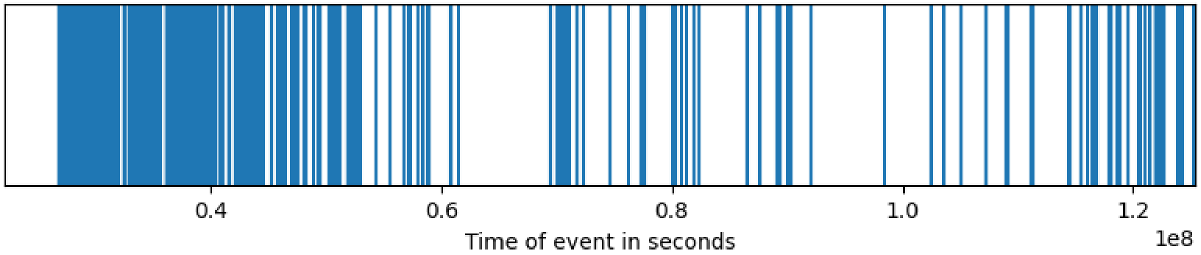
Aversion to form too many ties
→ probability to form new ties should decrease with in-degree k_{in}.

Model from (Ubaldi et al., 2016)

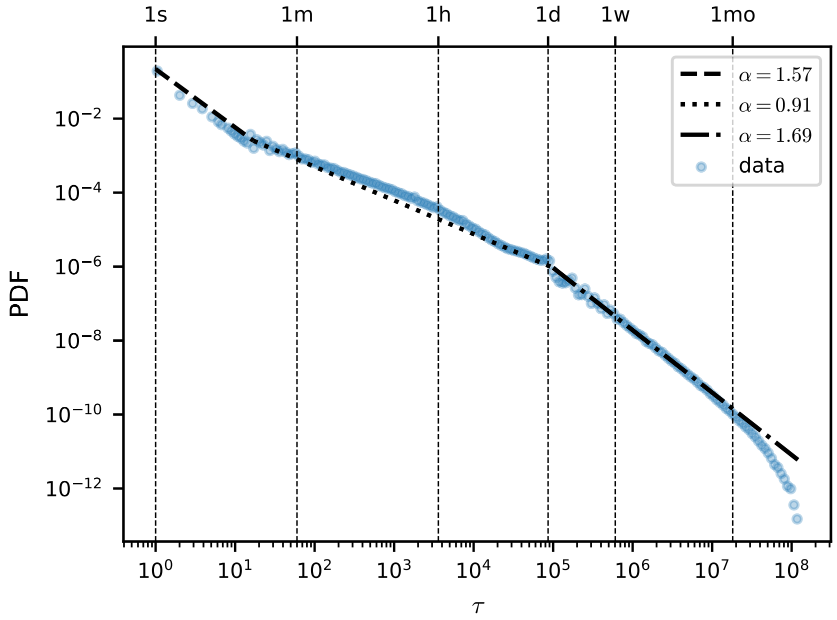
$$p_{\text{new}}(k_{\text{in}}) = \left(1 + \frac{k_{\text{in}}}{c}\right)^{-b}$$



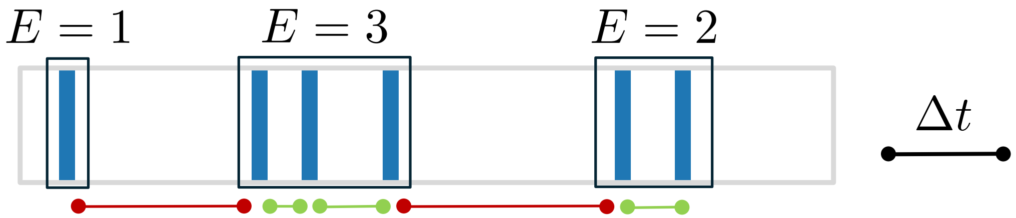
For all channels, get times between two forwarded messages = *inter-event times* τ



↪ $f(\tau)$ is piecewise power-law, with two main regimes separated by $\tau = 1 \text{ day}$



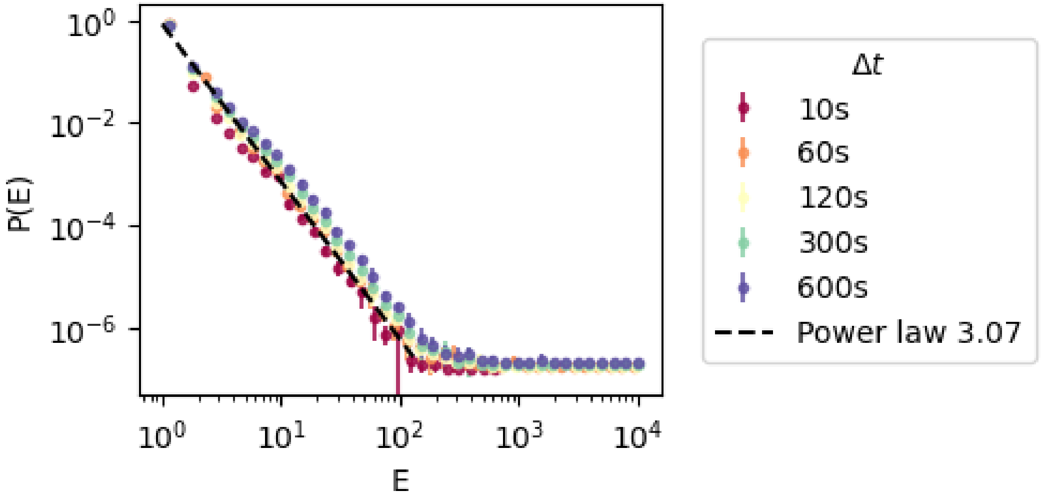
Investigate shape of distribution of *burst train sizes* E (Karsai et al., 2012):



We do have

$$p(E) \sim E^{-\beta}$$

↪ forwarding is bursty



Topology

- clustering
- power-law in/out-strength distributions
- language assortativity
- tendency to reinforce existing ties

Time

- two regimes
- burstiness

Topology

- clustering
- power-law in/out-strength distributions
- language assortativity
- tendency to reinforce existing ties

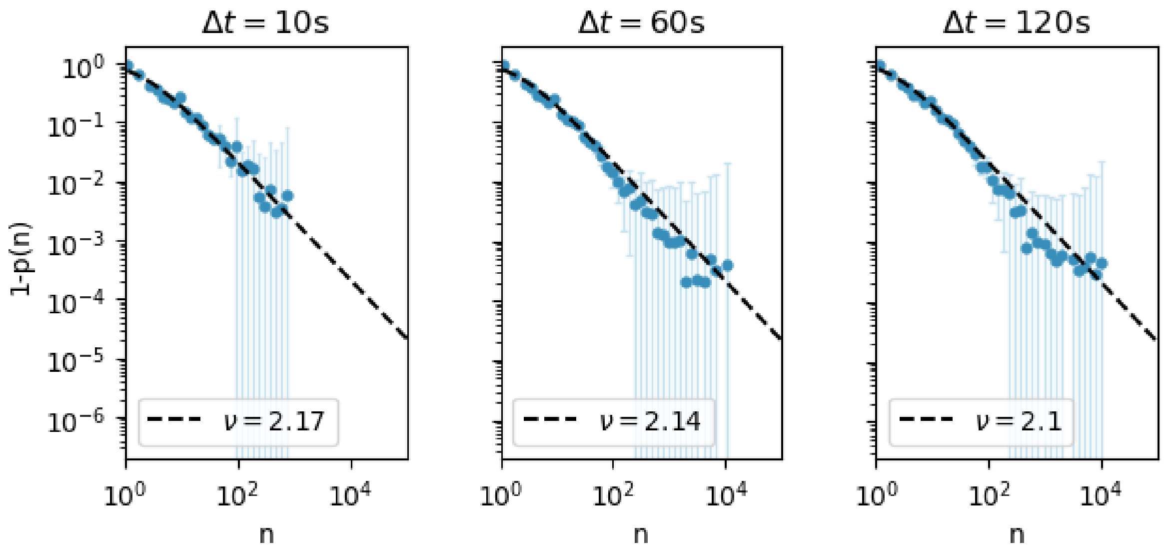
Time

- two regimes
- burstiness

Simple-enough model that can reproduce these features?

↪ Could help simulate contagion model or equivalent and test effect of interventions on synthetic networks

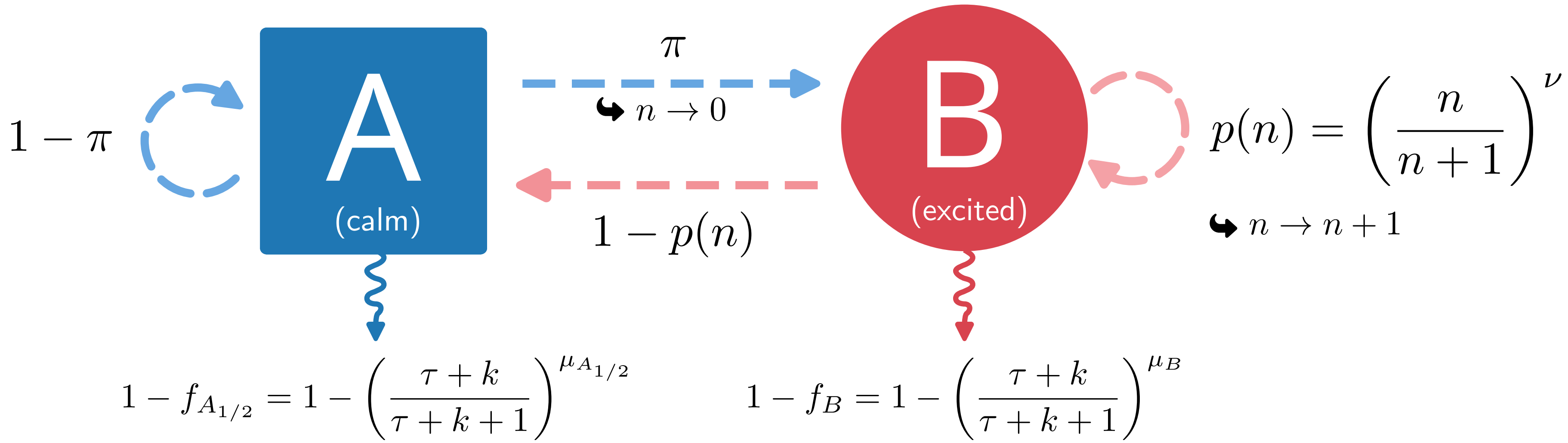
With already n events in a burst train, probability $p(n)$ to generate another within the same train?



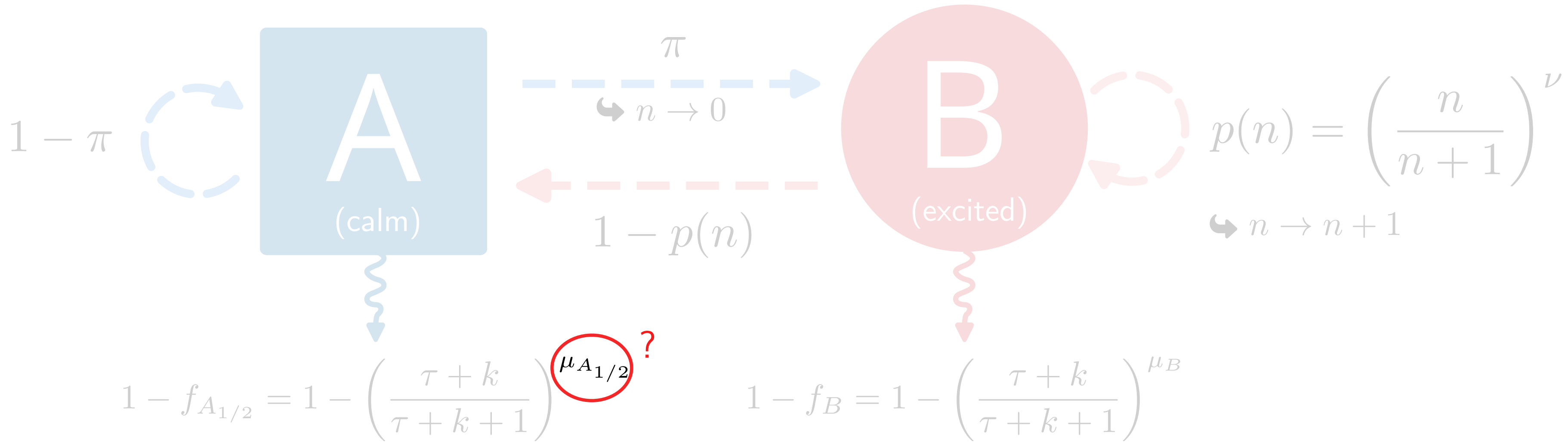
→ Train size distribution generated from memory process (Karsai et al., 2012)

$$p(E) \sim E^{-\beta} \Leftrightarrow p(n) = \left(\frac{n}{n+1}\right)^{\nu} \qquad \text{with } \nu \approx \beta - 1$$

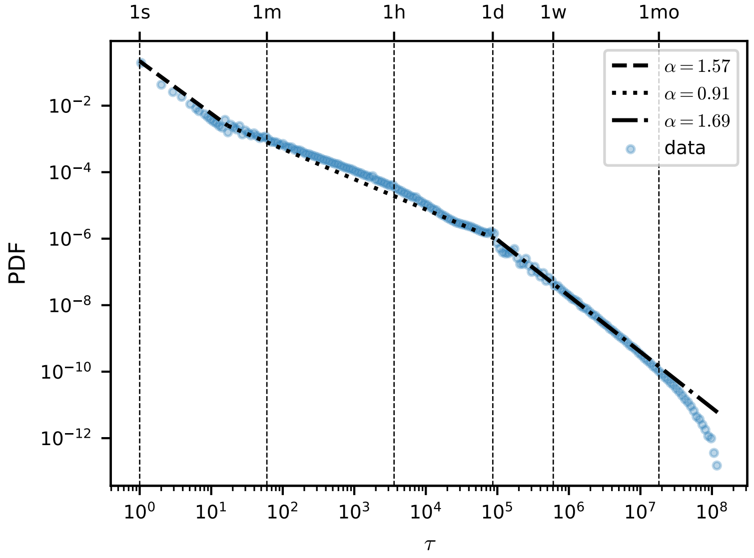
Generalisation from (Karsai et al., 2012)



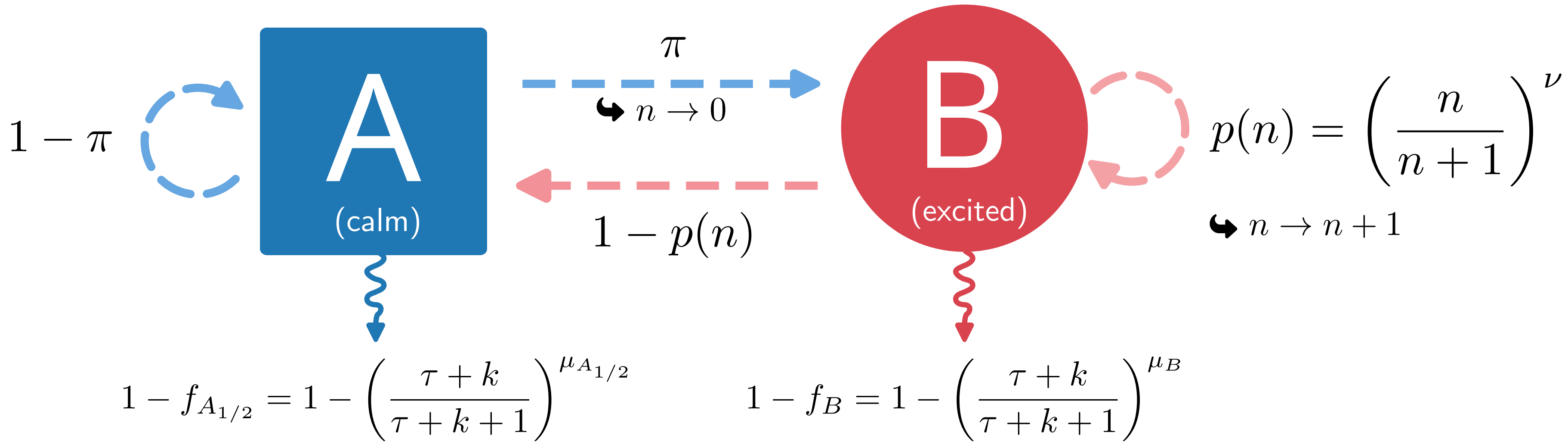
Generalisation from (Karsai et al., 2012)



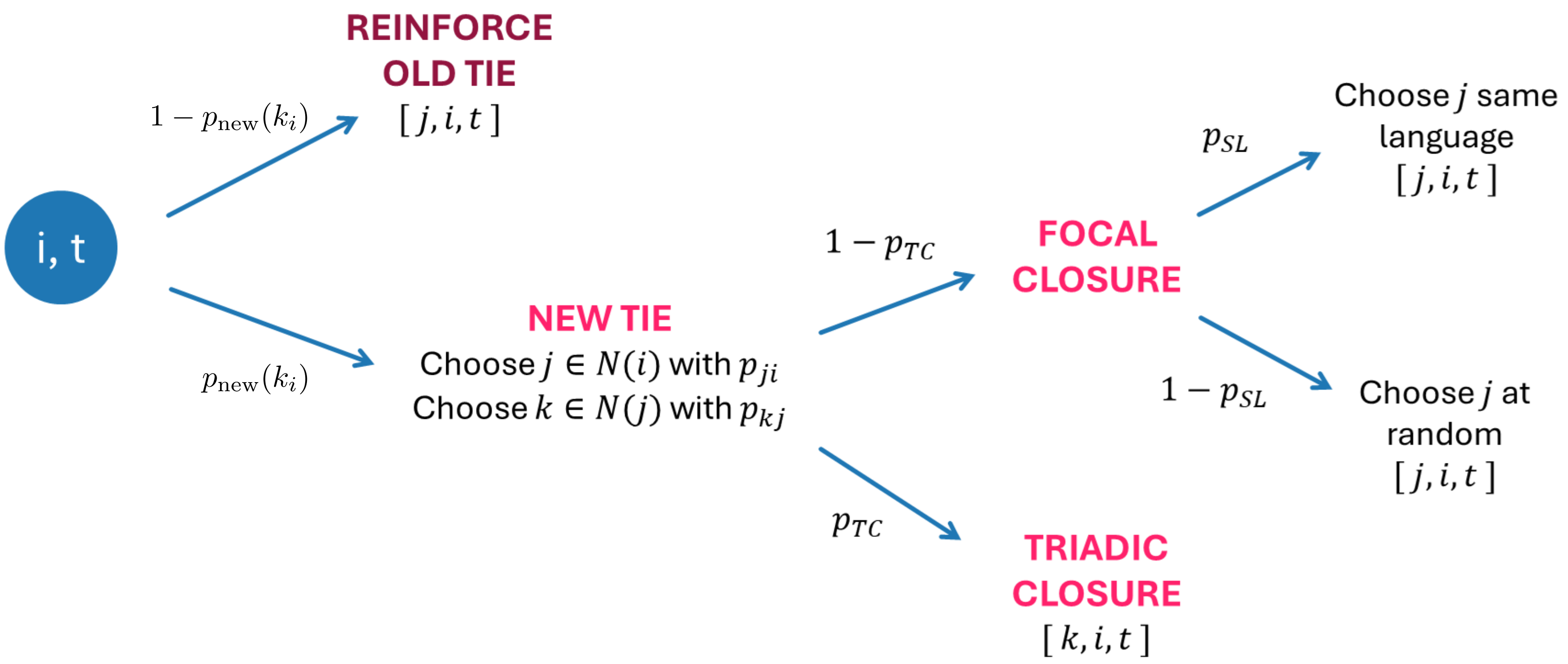
Generalisation from (Karsai et al., 2012)



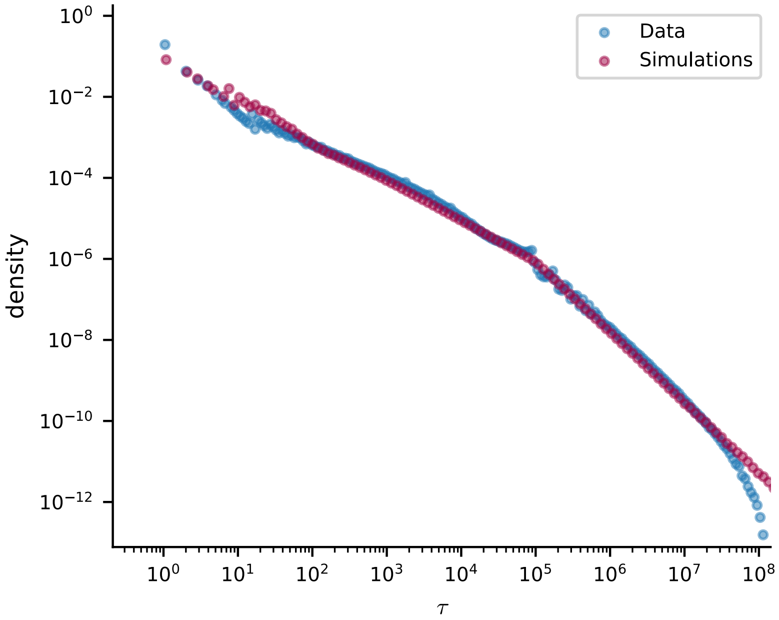
Generalisation from (Karsai et al., 2012)



Adapted from (Laurent et al., 2015)



Fitted time model (π, $\mu_{A_{1/2}}$, μ_B, k) to reproduce piecewise power-law $p(\tau)$



↪ *It fits* (+ it runs fast: $\sim 10s$) ($\pi \approx 0.20, \mu_{A_1} \approx 0.019, \mu_{A_2} \approx 0.74, \mu_B \approx 4.8, k = 81$)

Can generate synthetic networks by creating event sequence for each node, and then pick who they forward using topology model.

Can generate synthetic networks by creating event sequence for each node, and then pick who they forward using topology model.

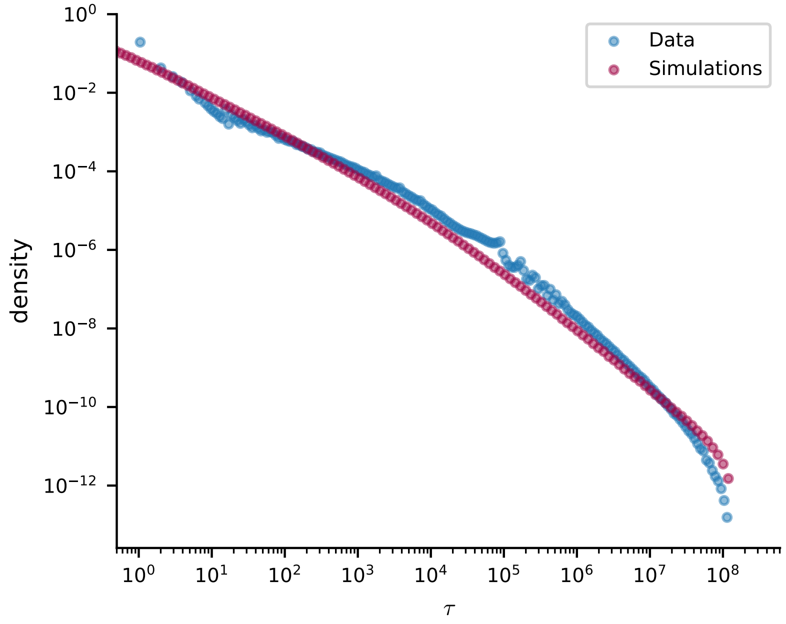
Issue: no guarantee average event rate for nodes is conserved.

Can generate synthetic networks by creating event sequence for each node, and then pick who they forward using topology model.

Issue: no guarantee average event rate for nodes is conserved.

→ What if we just contract/dilate time to fit event rates?

↪ slight deformation of $p(\tau)$



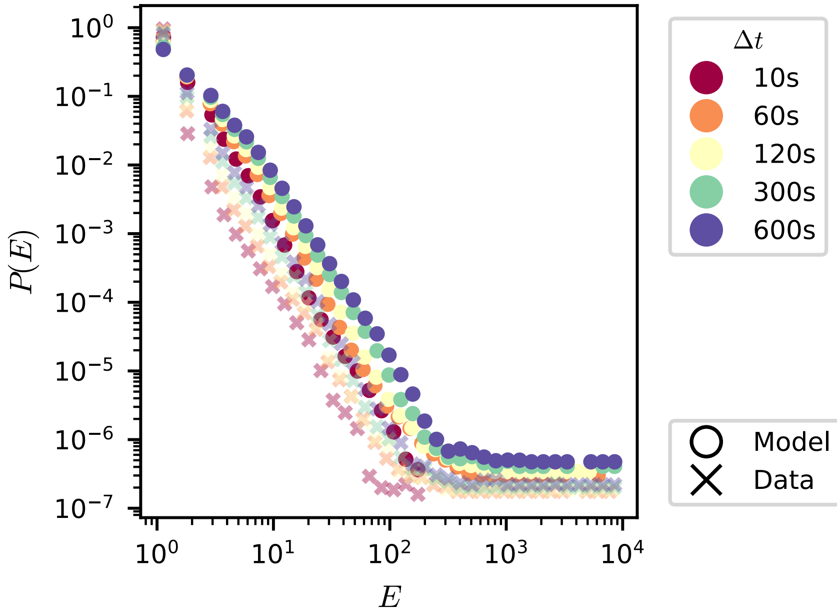
Can generate synthetic networks by creating event sequence for each node, and then pick who they forward using topology model.

Issue: no guarantee average event rate for nodes is conserved.

→ What if we just contract/dilate time to fit event rates?

↪ slight deformation of $p(\tau)$

↪ very similar β in $p(E) \sim E^{-\beta}$



What we've shown...

- Network of Telegram channels is very social-network-like
- Main mechanisms behind its emergence: tie reinforcement, clustering, language assortativity + memory process

...and what this leads to

- Model information propagation and effect of interventions
- Very global view of temporal process: what about local coordination?

and much more!

Thanks for your attention 🤗

@TLouf
@TLouf
tlouf@fbk.eu

Bibliography

Baumgartner, J., Zannettou, S., Squire, M., & Blackburn, J. (2020). The Pushshift Telegram Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 840–847. https://doi.org/10.1609/icwsm.v14i1.7348

Karsai, M., Kaski, K., Barabási, A.-L., & Kertész, J. (2012). Universal Features of Correlated Bursty Behaviour. *Scientific Reports*, *2*(1), 397. https://doi.org/10.1038/srep00397

Laurent, G., Saramäki, J., & Karsai, M. (2015). From Calls to Communities: A Model for Time-Varying Social Networks. *The European Physical Journal B*, *88*(11), 301. https://doi.org/10.1140/epjb/e2015-60481-x

Ubaldi, E., Perra, N., Karsai, M., Vezzani, A., Burioni, R., & Vespignani, A. (2016). Asymptotic Theory of Time-Varying Social Networks with Heterogeneous Activity

and Tie Allocation. *Scientific Reports*, *6*(1), 35724. https://doi.org/10.1038/srep
35724