Emerging Models and Paradigms in Network Science

Part #1: Temporal Networks

Ph.D. Course at University of Calabria

27–31 May 2024

Francesco Cauteruccio

DIEM, University of Salerno

fcauteruccio@unisa.it

francescocauteruccio.info

Studying Complex Systems

- Overview of large, integrated systems often requires zooming out from details.
- Systems such as the Internet, metabolism, proteome, and social networks can be represented as graphs.
- Networks often serve as the infrastructure for dynamical systems (e.g., data traffic, disease spread).
- Graph modeling helps estimate network behavior without studying actual dynamics.

Why graph modeling?

- Advantages of graph modeling:
  - Assessing influence between network parts.
  - Evaluating network optimization.
  - Identifying vertices with similar roles.
- Enhancing models with additional details:
  - Weighted networks include edge weights.
  - Spatial networks consider the position of vertices.

What about modeling a non-static scenario?

Introducing Temporal Networks

- Temporal networks incorporate the time dimension, showing when edges are active.
- Traditional network studies often aggregate contacts into edges, losing temporal details.
- Segmenting data into time windows still misses aspects of temporal structures.
- Temporal networks show non-transitive edge connections due to time ordering.
- The time ordering of connections affects network behavior beyond static network capabilities.

[References] The content in this presentation is mainly based on Holme, Petter, and Jari Saramäki. "Temporal networks." Physics reports 519.3 (2012): 97-125 [15].

Why studying temporal networks?

- Traditional network modeling separates static networks from dynamic systems.
- Temporal network approaches integrate timing information into the network structure.
- Suitable systems for temporal network modeling:
  - Information flow (e-mail, phone calls, social media).
  - Spreading dynamics of electronic and biological viruses.
  - Genetic regulation activation sequences.
  - Functional brain networks' time-domain features.
  - Evolving food webs and species networks influenced by environmental conditions.
  - Self-assembled networks of wireless devices and distributed computing systems.

Example of possible scenarios I

- Person-to-person communication [9, 25, 11]
  - Common examples are emails and mobile phone text messages, as well as instant messages in social networks.
  - Also not instantaneous interactions (e.g., phone calls) can be considered.
- One-to-many information dissemination [17, 16]
  - Contexts in which information is being broadcasted from a source to one or more target
  - Taking the time dimension into account enables a series of interesting analyses
- Proximity
  - Proximity patterns of human: *who is close to whom at what time*
    - Contacts between two patients in a hospital, movement of animals between farms, etc (§ 2.3)
  - Example projects: Reality Mining [8], SocioPatterns [5]

Example of possible scenarios II

- Cell biology
  - Different systems in cell and microbiology can be modeled as networks and naturally benefit from a temporal modeling
  - Interactome – set of molecular interactions in a cell [23]
    - Proteins or lighter molecules can connect to one another to perform biological functions
  - Metabolism network – set of chemical reactions that occur in a healthy organism [7]
    - Vertices are the molecular species that are connected if they are involved in the same chemical reaction
- Neural and brain networks [4, 10, 24]
  - Several levels of structural and temporal connectivity
    - Spiking patterns of individual neurons, functional connections between brain areas
  - Vertices are associated with time series, and depend on the experimental technique
    - Individual sensors for electroencephalography (EEG) and magnetoencephalography (MEG)
    - Voxels for functional magnetic resonance imaging (fMRI)
  - Edges are assigned if the signals are correlated/in phase

Example of possible scenarios III

- Artificial neural networks (ANN) [2]
  - The architecture of perceptrons and components of an ANN can be studied via network modeling
  - The time domain is understudied regarding this topic
- Ecological networks
- Distributed computing
- Infrastructural networks

Definition of temporal networks

- Generally, two definitions of temporal networks are employed in the literature
  - Contact sequence – a model indicating whether two vertices interact with each other at *certain times*
  - Interval graphs – the edges are not active over a set of times but rather over a set of *time intervals*

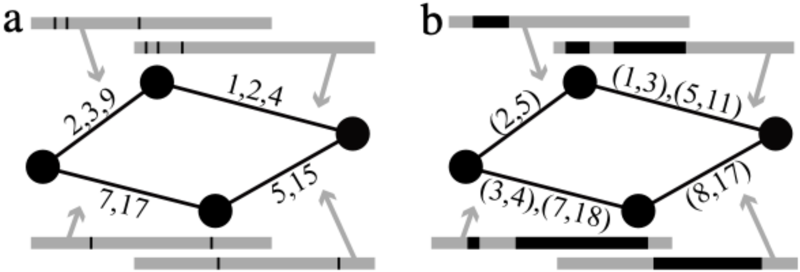

Figure: Contact sequences and interval graphs. Source: [15]

Contact sequence

- Defined over a set V of N vertices
- Interactions are represented via a *contact sequence*
  - A contact sequence C is a set of contacts,
  - each contact is a triple (i, j, t) where $i, j \in V$ and t denotes time
- Equivalently, we can represent the whole system by
  - the set V of vertices,
  - a set E of M edges,
  - for $e \in E$, a non-empty set of times of contacts $T_e = \{t_1, \ldots, t_n\}$.
- Typical systems suitable to be represented as a contact sequence:
  - Communication data,
  - Proximity data.
- Here, the duration of a contact is not considered.

Interval graphs

- Edges are active over a set of intervals $T_e = \{(t_1, t_1'), \dots, (t_n, t_n')\}$
  - t_k indicates the beginning of the interval, while t_k' indicates the end, $k = 1..n$
- Typical systems suitable to be represented as an interval graph:
  - Proximity networks (a contact represents two individuals that have been close for some extent of time)
  - Ecological networks (e.g., seasonal food webs)
  - Infrastructural systems (e.g., data network)

Restrictions

- Network theory poses some restrictions on what we deal with
  - E.g., in simple graphs an edge never occurs twice between the same nodes
- Assumption #1 – (contact sequence) A triple of a contact sequence never occurs twice
- Assumption #2 – (interval graphs) Consider two intervals $(t_i, t_i'), (t_j, t_j') \in T_e$, then the following must hold
  - $t_i < t_i'$,
  - $t_j < t_j'$,
  - $t_i < t_j$ iff $t_i' < t_j$

Lossless and Lossy Representations I

- Holme [13] proposes two representations: lossless and lossy.
- Lossless representations – these carry all information about a temporal network
  - Contact sequences,
  - Graph sequences – aka multilayer networks,
  - Dynamic networks – static networks evolving over time with respect to an underlying mechanism (e.g., preferential attachment [3])
  - Time-node graphs – each node is represented for every timestep
  - Time series of contacts on a static graph
  - Time-lines of contacts
  - Adjacency tensors

Lossless and Lossy Representations II

- Lossy representations – representations where some information of the original temporal network is lost
  - Weighted graphs – edge weights count the number of contacts between two nodes
  - Reachability and influence graphs
  - Time-window graphs – a static network including links present in the temporal network within a time window
  - Concurrency graphs – two nodes are linked if they had contacts both before a certain t_{start} and after a certain t_{stop}
  - Difference graphs – a static network highlighting the change over time (i.e., the difference between two consecutive timesteps)

Examples

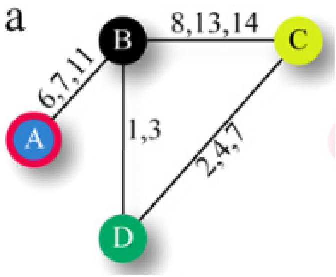

Figure: A graph visualization of a contact sequence. Source: [15]

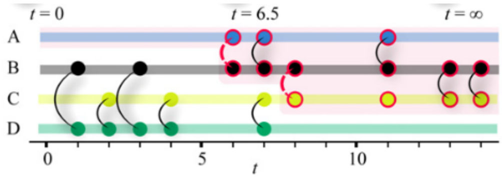

Figure: A timeline visualization of a contact sequence. Source: [15]

Measures

- The topological structure of static networks can be characterized by a plethora of measures
- When the additional degree of freedom of time is include, many of them need rethinking or revising
- We discuss the following aspects for measures
  - Time-respecting paths and reachability,
  - Distances, latencies, and fastest paths,
  - Connectivity and components,
  - Diameter and network efficiency,
  - Centrality measures,
  - Patterns and motifs,
  - Burstiness

Time-respecting paths I

- In a static graph, a path is a sequence of edges such that one edge ends at the node where the next edge of the path begins
- In temporal networks, paths are defined as sequences of contacts with non-decreasing times that connect sets of vertices
  - Such paths are called time-respecting paths (TRP) [14]
  - For instance, in Figure 2, there are TRPs from A to C (e.g., $(A, B, 7)$, $(B, C, 8)$).
- TRPs are not transitive
  - A TRP from i to j via k does not imply that there is a path from i to k
    - A path from k to k via j exists only if the first contact on the $j - k$ path takes place *after* the last contact on the $i - j$ path.

Time-respecting paths II

- TRPs define which nodes can be reached from which other nodes within some **observation window** $t \in [t_0, T]$
- The set of nodes that can be reached by TRPs from i is called the **set of influence of** i.
  - The **reachability ratio** is the average fraction of nodes in the sets of influence of all nodes.
- The **source set** of i is the set of nodes that can reach i through TRPs within a given observation window t
  - In spreading dynamics, this set consists of all nodes that can have been the source of an infection infecting i.
- Both sets are time-dependent, therefore an often useful analysis is to monitor them as a function of time
  - Study how many nodes may reach vertex i by TRPs by time t', when the paths begin no earlier than $t < t'$

Distances, latencies, and fastest paths I

- *How quickly vertices can reach each other through TRPs?*
- **Duration of a TRP** – the difference between the last and first contacts on the path
  - Also called temporal path length
- **Fastest TRP between two nodes** – the shortest TRP between two nodes
  - The shortest time within which i can reach j is called their **latency**
- In literature, the term distance is used to measure the number of links, while duration/latency is for measuring times
- Latency is a useful measure in several scenarios
  - Distributed computing – keeping track of the age of information that a node has about other nodes
  - Information spreading – the fastest possible trajectories of information between nodes

Distances, latencies, and fastest paths II

- Consider a node i at time t in a temporal network
- We denote with $\phi_{i,t}(j)$ the latest time before t such that information from j can have reached i by time t
  - $\phi_{i,t}(j)$ is called quantity i's *view* of j's information at time t
- Also, $\lambda_{i,t}(j) = t - \phi_{i,t}(j)$ is called j's *latency* with respect to i at time t
  - It is a measure of how old i's information coming from j is at time t
- The vector $[\phi_{i,t}(1), \ldots, \phi_{i,t}(N)]$ is called i's *vector clock*
  - It captures the partial ordering of events in a distributed system for the node i
  - Example: `examples/temporal-network-0.md`
- Usually, these measures are used to characterize the overall "velocity" of the network
  - Measuring how quickly vertices can, on average, transmit something.
  - How can we do it?

Distances, latencies, and fastest paths III

- An approach is to enumerate all fastest TRPs between vertices and then compute the average duration
  - However, this does not take into account the frequency of the paths
- Computing average latency is also difficult
  - Close to the end of the observation window, latency becomes infinite (paths no longer have enough time to be completed)
  - Some posed constraints on paths or just repeated values considered in previous steps [19]
- In conclusion, there is no de facto approach for averaging such quantities.

Diameter and Network Efficiency I

- For a static network $G(V, E)$
  - Diameter: $D = \max_{u,v \in V} d_{u,v}$, i.e., the largest distance between any pairs of vertices,
  - (Global) Efficiency: $E = \frac{1}{|V|(|V|-1)} \sum_{u \neq v \in V} \frac{1}{d_{u,v}}$ the avg over inverse path lengths of all paths
- No general definitions in temporal networks
  - The diameter could be defined as the longest average latency,
  - [6] requires that the diameter should be a number as small as possible such that increasing it would not make you find more pairs of vertices connected by a TRP that is shorter than the diameter
  - [20] defines network efficiency as the harmonic average of the latency
    - It combines the average latency and the reachability ratio defined in [12]

Centrality measures I

- Used for identifying important vertices w.r.t. some properties,
  - Degree centrality, Betweenness centrality, PageRank centrality, etc
- There is usually no unique candidate for the temporal version of a static centrality measure
- Simple approach: replace the role of paths by TRP
  - Closeness centrality C_C in temporal networks [22]: $C_C(i,t) = \frac{N-1}{\sum_{j \neq i} \lambda_{i,t}(j)}$, where $\lambda_{i,t}(j)$ is the latency between i and j.
  - Efficiency-based closeness centrality $C_E(i,t) = \frac{1}{N-1} \sum_{j \neq i} \frac{1}{\lambda_{i,t}(j)}$.
  - Betweenness centrality C_B measures the fraction of shortest paths passing through a given node
    - It is defined in static networks as $C_B(i) = \frac{\sum_{i \neq j \neq k} \nu_i(j,k)}{\sum_{i \neq j \neq k} \nu(j,k)}$, where $\nu_i(j,k)$ is the number of shortest paths between j and k that pass i, and $\nu(j,k)$ is the total number of shortest paths between j and k.
    - It is generalizable to temporal networks by adding a dependence on time t and counting the fraction of fastest TRPs that pass through the given node [21]

Centrality measures II

- Matrix-based centrality measures like *eigenvector centrality* and *PageRank* are less straightforward to be generalized to temporal networks
  - Any generalization would have to use 3D tensors representing the temporal networks
- A simple algorithms to generalize the eigenvector centrality
  1. Start with a centrality value 1 at each node,
  2. At every contact between nodes i and j, let C_E values after the contact at timestep t be

     $$C_E^{t+1}(i) = \rho C_E^t(i) + (1-\rho)C_E^t(j)$$

     and

     $$C_E^{t+1}(j) = \rho C_E^t(j) + (1-\rho)C_E^t(i)$$

     where the parameter $\rho \in \mathbb{R}_{>0}$ sets the rate of centrality transmitted at a contact (if ρ is large it puts a bigger emphasis on recent contacts.

Patterns and Motifs

- Also in this case, there is usually no unique candidate for the definition of patterns and motifs in temporal networks
- Data mining-inspired approaches: *subgraph mining* [18]
  - The support set $S(G')$ of a subgraph G' is the set of timesteps when $G' \subseteq G_t$
    - G_t is the graph of all edges active at time t in a temporal network
  - Persistent subgraphs are graphs having support larger than a certain threshold.

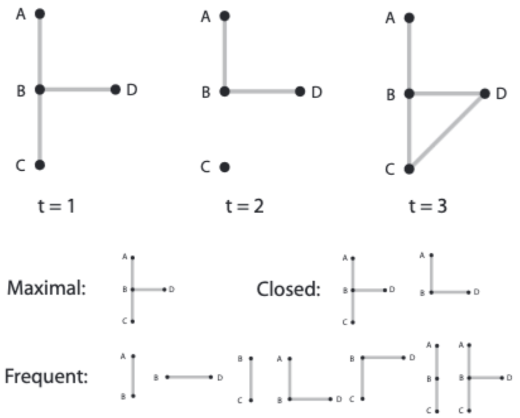

Figure: A temporal network with three timesteps and various types of frequent subgraphs at minimum support 2. Source: [18]

Patterns and Motifs

- A network motif [1] is an equivalence class of subgraphs that is overrepresented in terms of its cardinality w.r.t. some null model in a network
  - i.e., a larger number of such subgraphs can be found than in a randomized reference system.
- Motifs are generally defined a priori or selected from snapshots of the network
  - A simple approach is to aggregate edges over a period of time, and count the different subgraphs in the snapshots
  - Also, motifs can be defined on the basis of static subgraphs, and their presence then assessed on the temporal version
  - For instance, [26] define motifs such as *chains*, *stars*, and *"ping pong"* subgraphs.
- Several aspects are still understudied
  - What reference or null models should be used?
  - Recurrence of temporal subgraphs – exactly the same or similar?

Representing Temporal Data as Static Graph

- The literature on static networks is abundantly larger than that on temporal ones
  - Analyzing static graphs is easier than analyzing temporal ones
- One approach to analyzing temporal networks is to derive static networks that capture both temporal and topological properties of the system
- The straightforward way: accumulate the contacts over some time to form edges
  - This creates a “projection” of the temporal network
  - While it is useful whenever the topological aspects are considered, it can discard information on the temporal ones
- There are different ways of encoding the temporal network structure into a static graph

Reachability Graphs

- Aka *path graphs* or *associated influence digraph*
- A static network having the same nodes as the temporal one is created
- Then, a directed edge from node i to j is created if there is a TRP from i to j in the temporal network
- This is very useful when dynamics on the network are considered
  - For instance, the average degree k of a reachability graph is the average worst-case outbreak size, which effectively tells how vertices can possibly affect which others

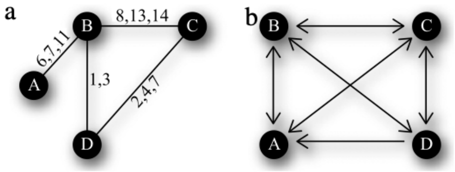

Figure: (a) shows a contact sequence, and (b) shows its reachability graph. Source: [15]

Line Graphs

- A line graph of a static graph G is a graph whose nodes are the edges of G that are connected if they share a vertex in G
  - Aka *interchange graph* or *dual graph*
- It finds several uses in epidemiology
  - The line graph is closely related to the structure of concurrent partnerships in the original graph
  - Also used extensively in the analysis of higher-order interactions

Transmission Graphs I

- A version of line graphs with temporal information [92]
- Built over an interval graph where an edge e is active over an interval $[t_{start}(e), t_{end}(e)]$
- The transmission graph has a directed edge from e to e' if e and e' share a node in the interval graph,

  $$t_{start}(e) < t_{start}(e') + \delta$$

  and

  $$t_{start}(e') < t_{start}(e)$$

  - where δ is used to identify the last possible time a member of an edge can transmit information.
- A transmission graph encodes the directionality arising from the order of non-concurrent relationships.

Transmission Graphs II

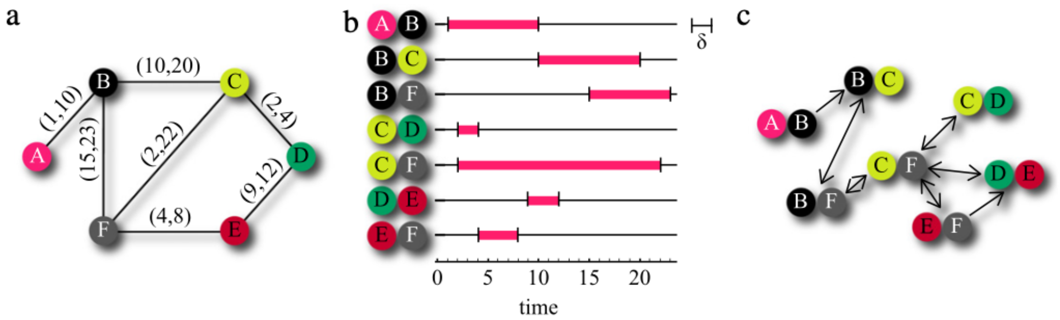

Figure: Transmission graphs. (a) shows an interval graph representation of a temporal network (where each edge has only one interval). (b) gives slightly reduced picture of the system, where a line corresponds to an edge in (a) and the active interval is indicated. The derived transmission graph is illustrated in (c). Source: [15]

Exercises

- Let's try playing with temporal networks
  - The exercises must be done in Python using iPython Notebook.
- Two steps
  1. Define a `TemporalNetwork` class
     - It represents a temporal network using the contact sequence representation.
     - It should at least provide methods to *(i)* add and remove contacts, *(ii)* compute and return all TRPs between two nodes i and j
  2. Use it to compute the *reachability ratio* and the *average latency*.
- A notebook to start with is available at `exercises/temporal-networks-0.ipynb`

Thanks for your attention!

Francesco Cauteruccio, Ph.D.
f.cauteruccio@unisa.it
https://francescocauteruccio.info

References I

Uri Alon.
Network motifs: theory and experimental approaches.
Nature Reviews Genetics, 8(6):450–461, 2007.

Alessia Amelio, Gianluca Bonifazi, Francesco Cauteruccio, Enrico Corradini, Michele Marchetti, Domenico Ursino, and Luca Virgili.
Representation and compression of residual neural networks through a multilayer network based approach.
Expert Systems with Applications, 215:119391, 2023.

Albert-László Barabási and Réka Albert.
Emergence of scaling in random networks.
science, 286(5439):509–512, 1999.

Ed Bullmore and Olaf Sporns.
Complex brain networks: graph theoretical analysis of structural and functional systems.
Nature reviews neuroscience, 10(3):186–198, 2009.

References II

Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani.
Dynamics of person-to-person interactions from distributed rfid sensor networks.
PloS one, 5(7):e11596, 2010.

Augustin Chaintreau, Abderrahmen Mtibaa, Laurent Massoulie, and Christophe Diot.
The diameter of opportunistic mobile networks.
In *Proceedings of the 2007 ACM CoNEXT conference*, pages 1–12, 2007.

Gal Chechik, Eugene Oh, Oliver Rando, Jonathan Weissman, Aviv Regev, and Daphne Koller.
Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network.
Nature biotechnology, 26(11):1251–1259, 2008.

Nathan Eagle and Alex Pentland.
Reality mining: sensing complex social systems.
Personal and ubiquitous computing, 10:255–268, 2006.

References III

Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi.
Entropy of dialogues creates coherent structures in e-mail traffic.
Proceedings of the National Academy of Sciences, 101(40):14333–14337, 2004.

F De Vico Fallani, Vito Latora, Laura Astolfi, Febo Cincotti, Donatella Mattia, Maria Grazia Marciani, Serenella Salinari, Alfredo Colosimo, and Fabrizio Babiloni.
Persistent patterns of interconnection in time-varying cortical networks estimated from high-resolution eeg recordings in humans during a simple motor act.
Journal of Physics A: Mathematical and Theoretical, 41(22):224014, 2008.

Petter Holme.
Network dynamics of ongoing social relationships.
Europhysics Letters, 64(3):427, 2003.

Petter Holme.
Network reachability of real-world contact sequences.
Physical Review E, 71(4):046119, 2005.

References IV

Petter Holme.
Modern temporal network theory: a colloquium.
The European Physical Journal B, 88:1–30, 2015.

Petter Holme, Christofer R Edling, and Fredrik Liljeros.
Structure and time evolution of an internet dating community.
Social Networks, 26(2):155–174, 2004.

Petter Holme and Jari Saramäki.
Temporal networks.
Physics reports, 519(3):97–125, 2012.

Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng.
Why we twitter: understanding microblogging usage and communities.
In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65, 2007.

References V

Ravi Kumar, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins.
On the bursty evolution of blogspace.
In *Proceedings of the 12th international conference on World Wide Web*, pages 568–576, 2003.

Mayank Lahiri and Tanya Y Berger-Wolf.
Structure prediction in temporal networks using frequent subgraphs.
In *2007 IEEE Symposium on computational intelligence and data mining*, pages 35–42. IEEE, 2007.

Raj Kumar Pan and Jari Saramäki.
Path lengths, correlations, and centrality in temporal networks.
Physical Review E, 84(1):016105, 2011.

John Tang, Mirco Musolesi, Cecilia Mascolo, and Vito Latora.
Temporal distance metrics for social network analysis.
In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 31–36, 2009.

References VI

John Tang, Mirco Musolesi, Cecilia Mascolo, Vito Latora, and Vincenzo Nicosia.
Analysing information flows and key mediators through temporal centrality metrics.
In *Proceedings of the 3rd Workshop on Social Network Systems*, pages 1–6, 2010.

John Tang, Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo, and Vito Latora.
Small-world behavior in time-varying graphs.
Physical Review E, 81(5):055101, 2010.

Ian W Taylor, Rune Linding, David Warde-Farley, Yongmei Liu, Catia Pesquita, Daniel Faria, Shelley Bull, Tony Pawson, Quaid Morris, and Jeffrey L Wrana.
Dynamic modularity in protein interaction networks predicts breast cancer outcome.
Nature biotechnology, 27(2):199–204, 2009.

Miguel Valencia, J Martinerie, Samuel Dupont, and M Chavez.
Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach.
Physical Review E, 77(5):050905, 2008.

References VII

Ye Wu, Changsong Zhou, Jinghua Xiao, Jürgen Kurths, and Hans Joachim Schellnhuber.
Evidence for a bimodal distribution in human communication.
Proceedings of the national academy of sciences, 107(44):18803–18808, 2010.

Qiankun Zhao, Yuan Tian, Qi He, Nuria Oliver, Ruoming Jin, and Wang-Chien Lee.
Communication motifs: a tool to characterize social communications.
In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1645–1648, 2010.