

Chapter 8

Metrics for Temporal Text Networks



Davide Vega and Matteo Magnani

Abstract Human communication, either online or offline, is characterized by when information is shared from one actor to the other and by what specific information is exchanged. Using text as a way to represent the exchanged information, we can represent human communication systems with a temporal text network model where actors and messages coexist in a dynamic multilayer network. In this model, actors and messages are represented in separate layers, connected by inter-layer temporal edges representing the communication acts—who and when communicate what information. In this chapter we revisit some measures specifically developed for temporal networks, and extend them to the case of temporal text networks. In particular, we focus on defining measures relevant for the analysis of information propagation, including the concepts of walk, path, temporal precedence and path distance measures. We conclude by discussing how to use the proposed measures in practice by conducting a comparative analysis in a sample communication network based on Twitter mentions.

Keywords Human communication · Communication system · Information · Temporal text network · Multilayer · Text · Distance measure · Precedence · Path

8.1 Introduction

The concept of communication is fundamental to study modern and contemporary societies (Luhmann 1995), and it is particularly important in social network analysis: many of the existing network-based models of social systems directly or indirectly represent communication processes. For example, if we focus on temporal social networks, empirical studies include conversations on social media (Magnani et al. 2012; Wang et al. 2021; Mathew et al. 2019), mobile telephone calls (Karsai et al.

D. Vega · M. Magnani (✉)
InfoLab, Department of Information Technology, Uppsala University, Uppsala, Sweden
e-mail: matteo.magnani@it.uu.se

D. Vega
e-mail: davide.vega@it.uu.se

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
P. Holme and J. Saramäki (eds.), *Temporal Network Theory*, Computational Social Sciences, https://doi.org/10.1007/978-3-031-30399-9_8

2011), and face-to-face interactions (Stehlé et al. 2011; Sapiezynski et al. 2019). Even when we consider static models of social networks, such as friendship graphs without any associated temporal information, many of the metrics used to analyze them are still based on the assumption that some information is shared through the network. For example, we can measure the ability of actors or groups of actors to efficiently spread information (closeness, diameter, Page-Rank centrality), or we can identify actors with the ability to influence existing information flows (betweenness centrality). In summary, the most typical application of social network models and in particular temporal social networks is to study systems of communication.

Despite the central role of information in communication systems, the information exchanged through the social ties has often been neglected in network analysis. The most popular methods for the analysis of social networks are defined on simple graph models, only including actors and their relationships, and hence temporal network analysis methods often only rely on the additional availability of time annotations. Studies on information diffusion processes often acknowledge the importance of considering the actors propagating the information, the times of the propagation, and the content. In practice, however, the content (e.g., text posted online) is used only to define how actors are connected with each other based on, for example, the order of links between blog posts (Leskovec et al. 2007) and private messaging (Caetano et al. 2019), who re-shared the same content in social media (Gomez Rodriguez et al. 2010) or how actors interact with messages shared across multiple social networks (Salehi et al. 2015; Roth and Cointet 2010a). Tamine et al. (2016) use the concept of polyadic conversation, a model where chains of Twitter user interactions (replies, mentions and retweets) during a time interval are first grouped into conversation trees, and then aggregated into a static weighted graph of interactions between authors. This type of graph aggregation has recurrently appeared in the literature of network modeling (Aragón et al. 2017) and information retrieval (Magnani et al. 2012), but there is no consensus on either what is the best method to build such model (e.g., how to compute the length of a conversation in terms of time and/or tree's depth) or how the textual content affects the grouping of actors. These are important limitations, because studying communication networks without considering *what* is communicated can only allow a partial understanding of the underlying social system (Deri et al. 2018).

This chapter is based on a model for temporal text networks designed to enable a more accurate representation of human communication (Vega and Magnani 2018). Temporal text networks describe communication events among actors, including the actors exchanging information, the textual representation of this information and the times when the communication happens. While this model is still limited to textual information, text is a very common way to communicate (for example by email, or via social media posts) and can also be used to represent other forms of expression, for instance oral communication that can be translated to text either manually or semi-automatically through speech-to-text algorithms, and also images, that can be turned into a set of keywords describing them (Vadicamo et al. 2017; Magnani and Segerberg 2021).

While mathematically temporal text networks can be seen as extensions of temporal networks, which are themselves extensions of simple networks, there are two important differences that require the introduction of specific analysis methods. The most intuitive difference is of course the presence of text. An additional and more subtle difference lies in the semantics of the temporal annotations.

In the literature on temporal social networks the time on edges is typically used to indicate when an edge exists, e.g., that during that time the two actors are in contact and can exchange information. An implicit assumption in existing works is that information can be exchanged at any time when an edge is active, and that the exchange of information is instantaneous.

When we explicitly model communication networks, we should make a difference between edges representing the possibility of communicating and edges representing the actual production and consumption of information. In many cases the first type of edges exist between all actors; for example, we can always send an email to an existing email address. Therefore, in this chapter we focus on edges representing communication acts, that is, the actual exchange of information. These acts may have a non-negligible duration, therefore the time annotations in a temporal text network indicate when the transmission of a (text) message starts and when it finishes. Examples where this is important are messages exchanged through networks where the communication channel has a physical delay, and asynchronous communication such as by email and via social media, where the text is sent at some time but in general only received at a later time.

This different semantics of the temporal edges in temporal networks and in temporal text networks requires the re-definition of some central concepts, such as time-consistent paths, which in turn leads to the definition of new specific metrics.

Finally, it is worth mentioning that Natural Language Processing (NLP) methods such as sentiment analysis (O'Connor et al. 2023; Dodds and Danforth 2010) have been used in the past to study the evolution of tweets, songs, blogs, presidential speeches without requiring information about the underlying communication structure (who exchanges these data sources and how), using only data from time-annotated documents and time series information (Lavrenko et al. 2000). The temporal text network model does not only allow researchers to use NLP methods during the analysis, but it provides specific metrics to combine them with other measures from temporal networks.

This chapter introduces the concept of path in temporal text networks and various metrics to characterize them. In Sect. 8.2 we introduce the temporal text network model to encode communication networks. In Sect. 8.3 we introduce the concepts of walk and path in temporal text networks, and in Sect. 8.4 we define alternative ways of summarizing a path, based either on the times when the communication acts happen or on the text exchanged through a path. Finally, in Sect. 8.5 we conclude with an empirical comparison of some of the measures introduced in this chapter in a sample network formed by the Twitter interactions between Swedish politicians.

8.2 Representing Temporal Text Networks

From a mathematical point of view, a *temporal text network* (Vega and Magnani 2018) can be represented as a triple (G, x, t) where $G = (A, M, E)$ is a directed bipartite graph representing the communication network, $x : M \rightarrow X$ is a mapping between the messages in M and a set of sequences of characters (text) in X and $t : E \rightarrow T$ represents the time associated to each edge, where T is an ordered set of time annotations. Edge directionality indicates the flow of the communication: $(a_i, m_k) \in E$ indicates that actor a_i has produced text m_k , while $(m_k, a_j) \in E$ indicates that actor a_j is the recipient of message m_k . Actors with out-degree larger than 0 are information producers, actors with in-degree greater than 0 are information consumers, and actors with both positive in- and out-degrees are information prosumers. For the sake of readability, we will sometimes use a compact notation, e.g., (a, m, t, x) to indicate an edge $(a, m) \in E$ where $t(a, m) = t$ and $x(m) = x$.

Figure 8.1 describes a working example we will use during the remainder of this chapter, representing a temporal text network with $|A| = 8$ actors, $|M| = 6$ messages and $|E| = 15$ edges. It is important to observe that, in most cases, the edges to/from a message have different time attributes; the only restriction imposed by the model is that $(a_i, m), (m, a_j) \in E \Rightarrow t(a_i, m) \leq t(m, a_j)$. In other words, a message can be consumed at different times by each actor (e.g., different social media users can check their notifications at different times), but can never be received before it has been generated (e.g., a user cannot access information that has not been shared yet).

This simple model can be used to differentiate between so-called unicast (messages m_2 and m_3 in the figure) and multicast (messages m_1, m_4 and m_5) communication. The model can also be used to represent a variety of communication platforms such as email and Twitter mention networks, and can be easily extended adding edges between actors or between messages to represent additional relationships such as a follower/followee network. Unless we explicitly mention it, in the remainder of the chapter we will ignore these extensions.

A similar model to represent temporal interactions is the *contact sequence* (Holme and Saramäki 2012; Gauvin et al. 2013) model, which expresses temporal networks as

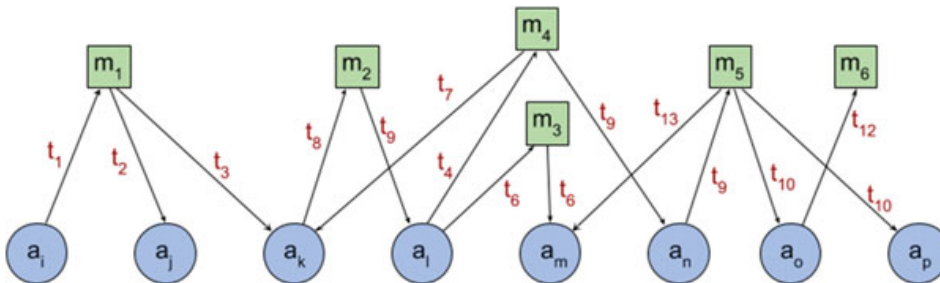


Fig. 8.1 A temporal text network model. Circles represent actors, squares represent messages and the edges between them represent the production and reception of the messages by the actors. Edges are also annotated with a time attribute $t_i \in T$

a set of directed edges (called contacts) during a finite span of time. While this model has been successfully used to study spreading processes of information (Lambiotte et al. 2013; Cheng et al. 2016; Caetano et al. 2019) or the structural evolution of social networks (Paranjape et al. 2017; Viard et al. 2016; Kim and Diesner 2017), the model ignores the role of the content of the messages.

A natural alternative to represent time in networks is to use a sequence of time-annotated graphs, forming a multilayer network (Dickison et al. 2016; Kivelä et al. 2014). In time-sliced models (Mucha and Porter 2010), for example, each one of the aggregated networks represents a fixed interval of time, and an edge e_{ij} exists in a slice if at least one contact has been registered between nodes i and j in the corresponding time interval. The aggregated graphs are sometimes weighted, in which case the edges have an assigned weight attribute w_{ij} proportional to the number of original edges, their frequency or another relevant time summarization function. In longitudinal networks, instead, the relations between the same or similar actors are detected at different points of time (Snijders 2005, 2014). From the modeling point of view there is not much difference between the two models, apart from the fact that in time-sliced networks the time intervals of two adjacent slices are typically contiguous, which is not necessarily true for longitudinal networks.

8.3 Path-Based Metrics

Metrics for simple networks are based on basic concepts in graph theory, such as adjacency and incidence, and on counting discrete objects such as edges. Temporal networks extend simple networks with time. This requires the extension of some basic concepts in graph theory, and as time is often represented as a real number or interval, then temporal measures also require some additional simple arithmetical operations, such as time difference.

Temporal text networks also contain a text attribute. Text is a much more complex type of data, with a large number of possible operations. For example, the comparison of two texts can be done using different models (edit distance, word overlapping, vector representation, etc.), applying different preprocessing operators (stemming, stop word removal, dictionary based word replacement) or mapping the text to other domains (for example sentiments or topics). While these choices are very important in practice, hard-coding all these details in the metrics would make the model very complex.

Therefore, as discussed by Vega and Magnani (2018), when dealing with temporal text networks we assume to have at least one of the following two types of text functions. The first type is used in a so-called *continuous analysis* approach, based on the idea of having different grades of similarity between messages. In this case we assume to have a distance function $d : M \times M \rightarrow [0, \infty)$, indicating how similar two messages are; if $d = 0$, the two messages are considered indistinguishable (for example because they contain the same text), and higher values of d indicate that the two messages are less similar. Notice that one can then plug specific functions into

the model based on the text operations described above. An example of a message distance function is the cosine of the angle between vector representations of the two texts.

The second type of functions is used in a so-called *discrete analysis* approach, where each message is assigned to 0, 1 or more classes. For each class i we have a function $c_i : M \rightarrow \{0, 1\}$, which returns 1 if the message belongs to class i , 0 otherwise. One example is a topic modelling function with k topics, where $c_i(m) = 1$ if m belongs to topic i . Notice that starting from a discretization function we can also define a text distance function, for example based on how many common topics are shared between the two input messages.

8.3.1 Incidence and Adjacency

In digraphs two vertices are adjacent if there is an edge between them, and two edges are adjacent if the tail of the first is the head of the second. In temporal text networks two vertices are adjacent *at time t* if there is an edge between them *at that time*. The concept of adjacency has also been extended to edges (also known as events or contacts): an edge entering a vertex is adjacent to an edge leaving the same vertex at a later time. This enables the definition of Δt -adjacency between edges, which is satisfied when they are adjacent and the time between them is less or equal than Δt . Note that this terminology is not completely consistent with the one in digraphs, where only vertices can be adjacent.

Temporal text networks differ from the previous cases in two regards. First, we do not need to extend the concept of adjacency to edges: we have two types of vertices (actors and messages), so for example the concept of adjacency between edges in temporal networks corresponds to adjacency between messages. This also means that we can retain the concept of incident edges from the theory of digraphs. Second, the idea of filtering those pairs of vertices that are close enough in time can also be extended to actors. In summary, all the concepts discussed above can be reduced to the following definitions, where v_i, v_j can be either actors or messages, with u_k being a node of the other type.

Definition 1 (*Edge incidence*) Let $e_1 = (v_i, u_k, t_1)$ and $e_2 = (u_k, v_j, t_2)$ be two edges in a temporal text network. We say that e_1 is incident to e_2 if $t_1 \leq t_2$.

Definition 2 (*Adjacency*) Let $e_1 = (v_i, u_k, t_1)$ and $e_2 = (u_k, v_j, t_2)$ be two edges in a temporal text network. Then:

1. v_i is adjacent to u_k at time t_1 .
2. v_i is Δt -temporally adjacent to v_j if $t_2 - t_1 \leq \Delta t$.
3. v_i is Δx -textually adjacent to v_j if $v_i, v_j \in M$ and $d(v_i, v_j) \leq \Delta x$.

Notice that the definition of incidence and adjacency hold independently of the type of vertices (v_i, u_k and v_j) involved. If $v_i, v_j \in A$ are actors, then their temporal

adjacency is defined by the delay between the production and consumption of the message $u_k \in M$. We call an edge from an actor a to a message m a *producer edge* (e_p), while an edge from a message m to an actor a is called a *consumer edge* (e_c). If $v_i, v_j \in M$ are messages, then their temporal adjacency is defined by the delay between when the intermediate actor consumes (e.g., receives) the first message and the time when it produces (e.g., sends) the second. For example, the producer edge $e_4 = (a_l, m_4)$ in Fig. 8.1 is incident to the consumer edge $e_{10} = (m_4, a_n)$, therefore actor a_l is Δt -adjacent to actor a_n for all $\Delta t \geq t_9 - t_4$.

8.3.2 Walks and Paths

Definition 3 (*Walk*) A *walk* in a temporal text network (also called a temporal walk) is a sequence of edges e^1, e^2, \dots, e^l where e^i is incident to e^{i+1} for all i from 1 to $l - 1$.

In the following we will write $a \in w$ to indicate that a vertex (actor or message) is present in walk w .

Notice that the definition above does not constrain the starting and ending vertices of a path to be actors or messages. However, we will often be interested in walks starting from an actor, because every message has a single producer in the model used in this chapter.

Definition 4 (*Path*) A *path* in a temporal text network (also called a temporal path) is a walk where no vertex (message or actor) is traversed twice.

Each path establishes a precedence relation between actors indicating that the network allows a flow of information between them. Similarly, we have a precedence relation between messages indicating that the two messages can be part of the same flow of information.

Definition 5 (*Temporal precedence*) An actor a_i *temporally precedes* another actor a_j if there is a path from a_i to a_j . A message m_i *temporally precedes* another message m_j if there is a path from m_i to m_j .

Figure 8.2 represents the temporal text network of Fig. 8.1 as a temporal sequence of edges between actors and messages. In this example, $w_1 = [e_4, e_7, e_8, e_9]$ and $w_2 = [e_4, e_{10}, e_{11}, e_{12}, e_{14}]$ are two walks of 4 and 5 edges.¹ The second walk is also a path, starting at an actor and ending in a message m_6 , but the first walk is not a path because the last edge $e_9 = (m_2, a_l, t_9)$ visits for a second time the actor a_l . Finally, notice that in this example a_l precedes actor a_k in path $p_1 = [e_4, e_7]$ and vice-versa in path $p_2 = [e_8, e_9]$, while m_3 precedes m_6 but not otherwise.

In some cases we may want to consider only those paths with a limited delay and with a limited textual difference between adjacent messages. We can thus use

¹ To simplify the notation, in this chapter we are assuming that $i \leq j \Rightarrow t_i \leq t_j$.

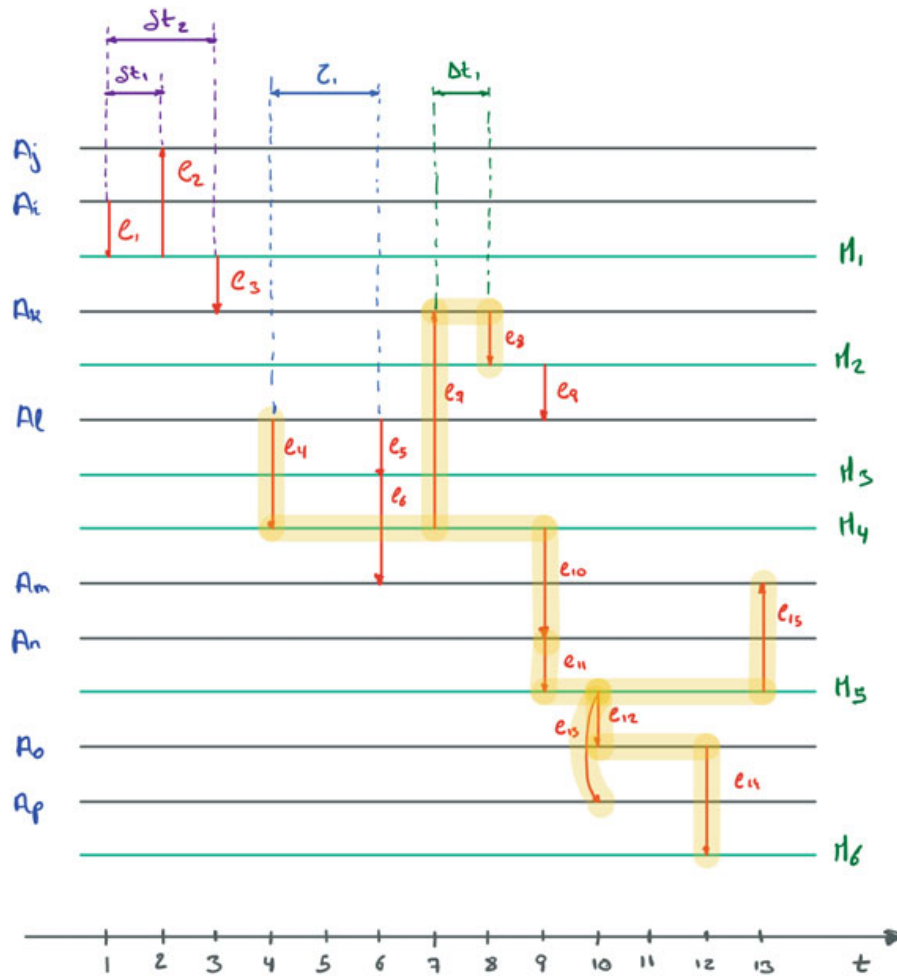


Fig. 8.2 Temporal text network represented as a sequence of edges. The horizontal lines represent the actors (gray color) and messages (green color) and vertical lines represent the transmission or consumption of a message. The shaded lines indicate all existing paths beginning at actor a_i at the exact time $t = 4$

the definitions of Δ -adjacency introduced above to select specific paths where sufficiently similar messages are exchanged often enough with respect to some user-defined thresholds.

8.4 Path Lengths

From now on we will focus on paths starting at an actor and ending at an actor. While a path can also start or end at a message, paths from and to actors are the ones providing the most accurate description of an information flow, because for every message there must always be an actor producing it, and messages that are not consumed by anyone (as message m_6 in our example) do not correspond to any exchange of information.

The length of a path in a temporal text network can be defined based on topology, on time, and on text.

The *topological length* is an unambiguous measure in simple and temporal networks, which are only made of vertices and edges. In a temporal text network a path contains actors, edges and messages, and the definition of length that is compatible with the one used in temporal networks corresponds to the number of messages in the path. This is because when a temporal network is translated into a temporal text network every edge is transformed into a message.

The *temporal length*, instead, defines the overall duration of the communication and is computed as the difference between the time of the last consumer edge and the time of the first producer edge in the path.

The topological and temporal length measures we have just described can be used to characterize the several paths that traverse our graph. In Fig. 8.2 we have highlighted all the existing paths starting at actor a_l at exactly $t = 4$, including those ending in a message. For example if we compare the path $p_1 = [e_4, e_{10}, e_{11}, e_{13}]$ with the path $p_2 = [e_4, e_{10}, e_{11}, e_{15}]$ we can see that both have the same topological length of 2 messages. However, while both paths start at the same time $e_4 = (a_l, m_4, t_4)$, the time of the last consumer edge is different and so their temporal length: $t(e_{13}) = t(m_5, a_p, t_{10}) \leq t(e_{15}) = t(m_5, a_m, t_{15})$.

Interestingly, in temporal text networks the temporal length of a path measures two different types of delays. On the one hand it measures the *transmission time* (δt) as the difference between the time of the consumer edge $t(e_c)$ and the time when the content has been produced $t(e_p)$. On the other hand, it indicates the *idle time* (τ) of the actors involved in the communication between two consecutive edges.

Definition 6 (*Transmission time*) Let $e_1 = (a_i, m, t_1)$ and $e_2 = (m, a_j, t_2)$ two incident edges, with $m \in M$. Then the quantity $t_2 - t_1$ is called transmission time.

Definition 7 (*Idle time*) Let $e_1 = (m_i, a, t_1)$ and $e_2 = (a, m_j, t_2)$ two incident edges, with $a \in A$. Then the quantity $t_2 - t_1$ is called idle time.

Once one has defined transmission and idle times, one can also compute the sum of all transmission times in a path, the sum of all idle times in a path, and the ratio between these values and the temporal length of the path. Back to our previous example, we can observe that the total transmission time of the messages in the first path $\delta_1 = (t_9 - t_4) + (t_{10} - t_9) = 6$ is three units smaller than in the second path $\delta_2 = (t_9 - t_4) + (t_{13} - t_9) = 9$ while their idle time is the same $\tau_F = t_9 - t_4 = 0$; which explains why the first path has a smaller temporal length.

The last type of length concerns the textual content in the path. Every time a message is exchanged, this increases the temporal length of the corresponding amount of time. Similarly, every time a new text is included in the path, this increases the textual information in it.

Definition 8 (*Textual length*) Given a text distance function, the *textual length* of a communication path is defined as the sum of the distances between the texts of all pairs of adjacent messages in the path.

This definition quantifies the variations between adjacent messages. At the same time, it is possible that the texts of the message keep being updated when transmitted through the path, but never significantly deviate from the original message. In this case, an alternative definition of length can be used to compute the maximum distance between any pair of messages.

In the case of discrete text analysis, where each message can belong to some classes (for example topics), this idea of estimating how homogeneous the text is across the path can be computed using a classical measure of entropy, for example the Shannon index:

Definition 9 (Entropy) Let c_1, \dots, c_n be text discretization functions mapping text into one of n classes. Given a path p , we define $\rho_i(p) = \frac{\sum_{m \in p} c_i(x(m))}{M_p}$, where M_p is the number of messages in p . The textual entropy of path p is then defined as:

$$H(p) = - \sum_{i=1}^n \rho_i(p) \ln \rho_i(p) \quad (8.1)$$

According to this definition, if all messages that are part of a path belong to the same class (e.g., to the same topic), then the textual entropy will be 0, indicating a homogeneous path when we look at its text. Higher values of entropy would indicate that multiple classes (e.g., topics) are included in the path. This information can be useful in various analysis tasks, including the identification of information flows (when the same textual content is transferred through the network) or community detection, where one wants a community to be homogeneous not only with respect to the topology but also with the exchanged messages.

Once we decide which definition of length to use, this defines what the shortest paths between any pair of actors are, which implies that we can compute all the existing network measures based on shortest paths, including closeness centrality, betweenness centrality, eccentricity, diameter, etc. For the definitions of these metrics we refer the reader to any basic book on network analysis.

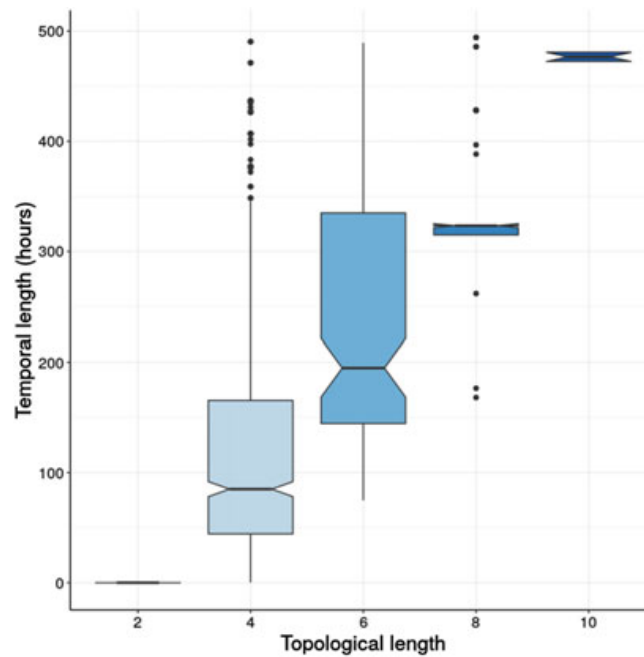
8.5 Empirical Study

In this section, we show an empirical comparison of the measures introduced in this chapter in a real communication network. Our sample dataset consists of all the public Twitter mentions (messages including another Twitter `@username`) written by Swedish politicians during January, 2019. The period of observation takes place four months after the Swedish general elections in 2018, and includes the time when the new government coalition was formed.² Our final network consists of $|A| = 886$ actors, including 26 politicians (8 information producers and 18 prosumers) and 860

² We considered only politicians who were either members of the parliament before the elections or were part of an electoral ballot.

Fig. 8.3 Temporal length.

Summary of the temporal length distribution for all shortest paths found in the Swedish politicians network, grouped by their topological path length. All topological paths involve an even number of hops because we are measuring only pairs of reachable actors

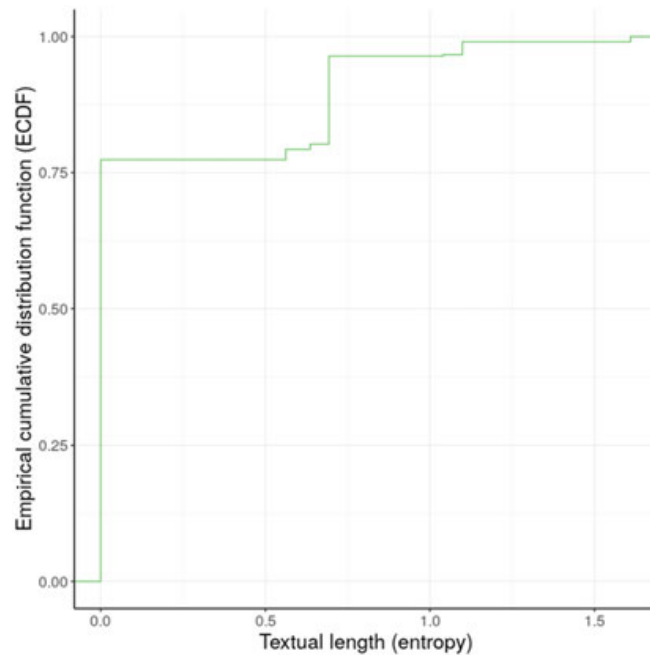


mentioned users (all of them consumers), $|M| = 1,707$ Twitter messages with their corresponding text and $|E| = 4,882$ edges between actors and messages. Modelling the reception time is more difficult, because many social media platforms like Twitter do not provide information about when and who consumed a piece of information. In our experiments we assumed that the consumption time of all messages is the same as the production time, which is not true in general (e.g., users are not always connected to all their social media and, even if they are, the tweet might be lost in the myriad of information provided by the user's wall).

Figure 8.3 shows, for each one of the 6,773 pairs of actors temporally reachable, a comparison of their topological and temporal shortest path length. It includes 5,787 (85.4%) paths with only two edges, representing two $\Delta 0$ -textually and temporally adjacent actors who have been in direct communication. The average temporal path length of the remaining paths increases with the number of hops (topological length) while its statistical dispersion is reduced, as we usually observe in other types of temporal networks (e.g., contact networks). For example, the 56 pairs of actors connected through 3 messages (6 hops) have an average communication time (shortest temporal length) of approximately 14 days. The order of magnitude of these numbers can be explained by the skewed distribution of roles (producer, consumer and prosumer) of the actors in the data and the small sample of the original social network.

Another important component to understand communication networks is the specific content their members intend to share with each other. For example, in a conversation within a group of close people the content (text) of the messages will be probably different between communications, while news spreading processes will probably have a more similar topic distribution. The consistency of the topics in an information cascade, therefore, can be a good metric to describe the dynamics of a complex system.

Fig. 8.4 Empirical cumulative distribution function (ECDF) of the textual length



Using the concepts described in Sect. 8.4, we have first identified the topics of the messages exchanged in our sample network and then, computed the textual length using the Shannon index described in Eq. 8.1 to identify the shortest paths of each pair of temporally reachable actors. While identifying the topics, we have used the hashtags as proxies, which is a simple and sometimes acceptable solution; but as we will see, problematic in practice. As we mentioned in the previous section, the definition of textual length assumes that there is a discretization function mapping the text *into at least one topic*. Hence, because many tweets do not contain any hashtag, their topic assignment is empty.

Figure 8.4 shows the empirical cumulative distribution function (ECDF) of the textual shortest path in our sample network. In this particular example, only 420 observations of 6,773 were computed, as many paths have an unidentified length, either because none of the messages have a topic assigned or because they contain only one message.

We can observe that more than 75% of the textual paths computed have 0 entropy, indicating that there is one single topic in the messages of the path. A closer look does not indicate any correlation of these results with the topological or temporal length of the paths. The minimum textual length paths include, for example, all the paths with 5 messages (10 hops) and 85.45% of the paths with 4 messages, but less than 50% of the paths with 3 messages.

8.6 Final Remarks

In this chapter, we have revisited some of the fundamental graph measures for temporal networks and extended them to be compatible with the temporal text network model for communication systems. We have shown that using the proposed model we can directly represent, in a simple but extensible way, all the elements necessary to study communication (time, text and topology), without requiring complex graph transformations. While mathematically temporal text networks are not much different from time-varying graphs, the semantics of its interactions and the presence of textual information in the model require the introduction of specific analysis methods. In particular, in this chapter we have focused on redefining the idea of connectivity and most of its related measures such as incidence, adjacency, paths and distance, providing alternative metrics for actors and messages when we found it was relevant and necessary. Finally, we have shown how the different distance measures can be used in practice to discover patterns of connectivity.

Temporal text networks can be seen as extensions of simpler network models, motivated by the need to capture more information that is not easily represented in the form of simple relationships between entities. This is just one example of a larger trend in network science, that has been present for a long time in social network analysis where there is often a clear need to obtain a non-binary understanding of social relations. Interdonato et al. review different types of features (beyond time and text) that have been considered in the literature (Interdonato et al. 2019), and temporal text networks are just one of the many approaches using complex network concepts to study text data (Oliva et al. 2021). Some of the approaches to study text using networks focus on the modelling of concepts or topics, that can be related to the discrete classes computed on temporal text networks to reduce the complexity of the individual text messages (Taskin et al. 2020; Camilleri and Miah 2021). These approaches are of particular relevance for temporal text network analysis when the concepts or topics can be related to actors (St-Ong et al. 2022) and the analysis has a temporal perspective (Roth and Cointet 2010a).

In Sect. 8.5 we have provided a simple empirical application of the presented measures, with the aim of exemplifying them. In the literature, networks including text and time have been used to study online political communication (Hanteer et al. 2018), online communities (Ustek-Spilda et al. 2021), and information spreading (Pereira 2021). While our example is about social media data, which is certainly a main field of application for temporal text networks, other types of data can also be modeled, including data from historical archives (Milonia and Mazzamurro 2022) and biomedical texts (Chai et al. 2020) but also non-textual data, where for example images can be discretized or compared using computer vision tools such as convolutional neural networks (Magnani and Segerberg 2021).

Each of the concepts and measures described in this chapter only focuses on some aspects of the data. As an example, a path is defined in a conservative way with respect to graph theory, not allowing the multiple appearance of the same actor. However, in some cases it can be useful to consider walks where the same actors

appear multiple times, as long as they exchange different messages. For example, as part of a longer discussion between two actors. Similarly, entropy and Δ -adjacency only consider respectively consecutive interactions in a path and the unordered set of all interactions, while in some cases we may be interested just in the difference between the first and the last message. In summary, these functions should be considered as a non-exhaustive set of fundamental building blocks, to be extended and expanded. Beyond their direct application to different analysis tasks, the basic measures described in this chapter can also be used to redefine other network measures so that they can capture more information from communication systems. Examples include centrality measures and community detection algorithms.

Acknowledgements We would like to thank Prof. Christian Rohner for his comments and suggestions.

This work was partially supported by the European Community through the project “*Values and ethics in Innovation for Responsible Technology in Europe*” (Virt-EU) funded under Horizon 2020 ICT-35-RIA call Enabling Responsible ICT-related Research and Innovation, and by eSSENCE, an e-Science collaboration funded as a strategic research area of Sweden.

References

- P. Aragón, V. Gómez, D. García, A. Kaltenbrunner, Generative models of online discussion threads: state of the art and research challenges. *J. Internet Serv. Appl.* **8**(1), 1–17 (2017)
- J.A. Caetano, G. Magno, M. Gonçalves, J. Almeida, H.T. Marques-Neto, V. Almeida, Characterizing attention cascades in whatsapp groups, in *Proceedings of the 10th ACM Conference on Web Science* (2019), pp. 27–36
- E. Camilleri, S.J. Miah, Evaluating latent content within unstructured text: an analytical methodology based on a temporal network of associated topics. *J. Big Data* **8**(1), 124 (2021)
- L.R. Chai, D. Zhou, D.S. Bassett, Evolution of semantic networks in biomedical texts. *J. Complex Netw.* **8**(1), cnz023 (2020). <https://doi.org/10.1093/comnet/cnz023>
- J. Cheng, L.A. Adamic, J.M. Kleinberg, J. Leskovec, Do cascades recur? in *Proceedings of the 25th International Conference on World Wide Web* (International WWW Conferences Steering Committee, 2016), pp. 671–681
- S. Deri, J. Rappaz, L.M. Aiello, D. Quercia, Coloring in the links: capturing social ties as they are perceived. *Proc. ACM Hum. Comput. Interact.* **2**(CSCW), 43:1–43:18 (2018)
- M. Dickison, M. Magnani, L. Rossi, *Multilayer Social Networks* (Cambridge University Press, 2016)
- P.S. Dodds, C.M. Danforth, Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *J. Happiness Stud.* **11**(4), 441–456 (2010)
- L. Gauvin, A. Panisson, C. Cattuto, A. Barrat, Activity clocks: spreading dynamics on temporal networks of human contact. *Sci. Rep.* **3**, 3099 (2013)
- M. Gomez Rodriguez, J. Leskovec, A. Krause, Inferring networks of diffusion and influence, in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10* (ACM, New York, NY, USA, 2010), pp. 1019–1028
- O. Hanteer, L. Rossi, D.V. D’Aurelio, M. Magnani, From interaction to participation: the role of the imagined audience in social media community detection and an application to political communication on twitter, in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2018), pp. 531–534
- P. Holme, J. Saramäki, Temporal networks. *Phys. Rep.* **519**(3), 97–125 (2012)

- R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron, A. Sala, Feature-rich networks: going beyond complex network topologies. *Appl. Netw. Sci.* **4**(1), 4 (2019)
- M. Karsai, M. Kivela, R.K. Pan, K. Kaski, J. Kertész, A.L. Barabási, J. Saramäki, Small but slow world: how network topology and burstiness slow down spreading. *Phys. Rev. E-Stat., Nonlinear, Soft Matter Phys.* **83**(2) (2011)
- J. Kim, J. Diesner, Over-time measurement of triadic closure in coauthorship networks. *Soc. Netw. Anal. Min.* **7**(1), 9 (2017)
- M. Kivela, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer networks. *J. Complex Netw.* **2**(3), 203–271 (2014)
- R. Lambiotte, L. Tabourier, J.C. Delvenne, Burstiness and spreading on temporal networks. *Eur. Phys. J. B* **86**(7), 320 (2013)
- V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, J. Allan, Mining of concurrent text and time series, in *SIGKDD Workshop on Text Mining* (2000), pp. 37–44
- J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in *International conference on Knowledge Discovery and Data Mining (KDD)* (2007), p. 420
- N. Luhmann, *Social Systems* (Stanford University Press, 1995)
- M. Magnani, D. Montesi, L. Rossi, Conversation retrieval from microblogging sites. *Inf. Retr. J.* **15**(3–4) (2012)
- M. Magnani, A. Segerberg, On the conditions for integrating deep learning into the study of visual politics, in *13th ACM Web Science Conference* (2021)
- B. Mathew, R. Dutt, P. Goyal, A. Mukherjee, Spread of hate speech in online social media, in *Proceedings of the 10th ACM Conference on Web Science, WebSci '19* (Association for Computing Machinery, New York, NY, USA, 2019), pp. 173–182
- S. Milonia, M. Mazzamurro, Temporal networks of ‘Contrafacta’ in the first three troubadour generations. *Digit. Sch. Humanities fqac018* (2022)
- P.J. Mucha, M.A. Porter, Communities in multislice voting networks. *Chaos: Interdiscip. J. Non-linear Sci.* **20**(4) (2010)
- B. O’Connor, R. Balasubramanyan, B.R. Routledge, N.A. Smith, From tweets to polls: linking text sentiment to public opinion time series, in *Proceedings of the Eleventh International Conference on Web and Social Media*, ed. by W.W. Cohen, S. Gosling (The AAAI Press)
- S.Z. Oliva, L. Oliveira-Ciabati, D.G. Dezebembro, M.S.A. Júnior, M. de Carvalho Silva, H.C. Pes-sotti, J.T. Polletini, Text structuring methods based on complex network: a systematic review. *Scientometrics* **126**(2), 1471–1493 (2021)
- A. Paranjape, A.R. Benson, J. Leskovec, Motifs in temporal networks, in *Proceedings of the 10th ACM International Conference on Web Search and Data Mining, WSDM '17* (ACM, New York, NY, USA, 2017), pp. 601–610
- F.S.F. Pereira, Caracterização da propagação de rumores no twitter utilizando redes textuais tempo-rais, in *Anais do Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)* (SBC, 2021), pp. 25–31
- C. Roth, J.P. Cointet, Social and semantic coevolution in knowledge networks. *Soc. Netw.* **32**(1), 16–29 (2010)
- M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, D. Montesi, Spreading processes in multilayer networks. *IEEE Trans. Netw. Sci. Eng.* **2**(2), 65–83 (2015)
- P. Sapiezynski, A. Stopczynski, D.D. Lassen, S. Lehmann, Interaction data from the copenhagen networks study. *Sci. Data* **6**(1), 1–10 (2019)
- T.A.B. Snijders, Models for longitudinal network data, in *Models and Methods in Social Network Analysis, Structural Analysis in the Social Sciences*, ed. by P.J. Carrington, J. Scott, S. Wasserman (Cambridge University Press, 2005), pp. 215–247
- T.A.B. Snijders, Siena: statistical modeling of longitudinal network data, in *Encyclopedia of Social Network Analysis and Mining* (Springer New York, New York, NY, 2014), pp. 1718–1725
- J. St-Onge, L. Renaud-Desjardins, P. Mongeau, J. Saint-Charles, Socio-semantic networks as mutu-alistic networks. *Sci. Rep.* **12**(1), 1889 (2022). Number: 1 Publisher: Nature Publishing Group

- J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.F. Pinton, P. Vanhems, High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One* **6**(8) (2011)
- L. Tamine, L. Soulier, L., Jabeur, F. Amblard, C. Hanachi, G. Hubert, C. Roth, Social media-based collaborative information access: analysis of online crisis-related twitter conversations, in *HT 2016 - Proceedings of the 27th ACM Conference on Hypertext and Social Media* (2016), pp. 159–168
- Y. Taskin, T. Hecking, H.U. Hoppe, ESA-T2N: a novel approach to network-text analysis, in *Complex Networks and Their Applications VIII, Studies in Computational Intelligence*. ed. by H. Cherifi, S. Gaito, J.F. Mendes, E. Moro, L.M. Rocha (Springer International Publishing, Cham, 2020), pp.129–139
- F. Ustek-Spilda, D. Vega, M. Magnani, L. Rossi, I. Shklovski, S. Lehuede, A. Powell, A twitter-based study of the European internet of things. *Inf. Syst. Front.* **23**(1), 135–149 (2021)
- L. Vadicamo, F. Carrara, A. Cimino, S. Cresci, F. Dell’Orletta, F. Falchi, M. Tesconi, Cross-media learning for image sentiment analysis in the wild, in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (2017), pp. 308–317
- D. Vega, M. Magnani, Foundations of temporal text networks. *Appl. Netw. Sci.* **3**(1), 26 (2018)
- T. Viard, M. Latapy, C. Magnien, Computing maximal cliques in link streams. *Theoret. Comput. Sci.* **609**(1), 245–252 (2016)
- L. Wang, A. Yang, K. Thorson, Serial participants of social media climate discussion as a community of practice: a longitudinal network analysis. *Inf., Commun. Soc.* **24**(7), 941–959 (2021)