LargeNetVis: Visual Exploration of Large Temporal Networks Based on Community Taxonomies – Supplemental material

Claudio D. G. Linhares*, Jean R. Ponciano*, Diogenes S. Pedro, Luis E. C. Rocha, Agma J. M. Traina, and Jorge Poco

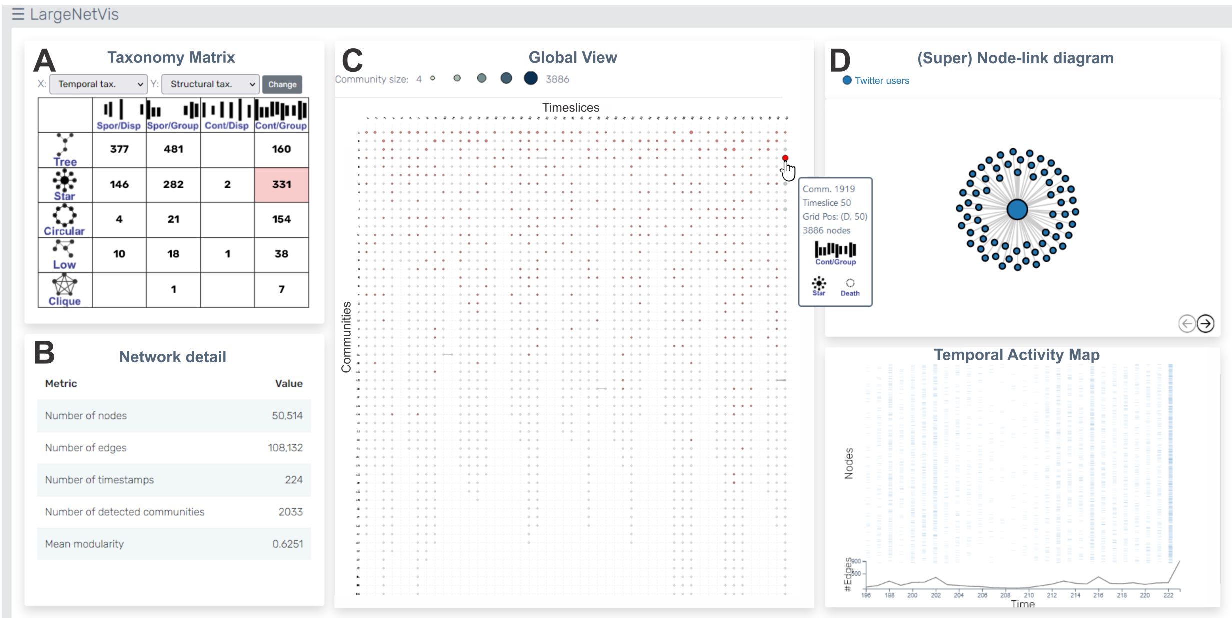

Fig. 1. Screenshot of LargeNetVis when analyzing the Twitter network. (A) There are 331 communities categorized as *Star* and *Continuous/Grouped* among (B) the 2,033 detected communities. (C) Global View exhibits all 2,033 communities while highlighting those with the above categorization of interest. (D) Our multilevel node-link diagram shows the most summarized representation of the largest community highlighted in (C), which is also the largest in the network, with 3,886 nodes and 5,035 edges.

A Link Length Reduction Algorithm

Algorithm 1 depicts the link length reduction strategy we use in Global View.

B Visual Scalability

Fig. 1 shows a screenshot of the LargeNetVis system when analyzing the Twitter network [1], which depicts retweets mentioning a famous Brazilian newspaper (*Folha de São Paulo*). There are 50,514 nodes (Twitter users) and 108,132 edges distributed in 224 timestamps, each comprising a 1-h interval. An edge is created when a user's tweet that mentions the newspaper is retweeted by another user — both users then become nodes in the network. When using 100 timeslices for analysis, the network is decomposed into 2,033 communities (Fig. 1(B)) that are simultaneously shown in Global View due to the default zoom-to-fit (Fig. 1(C)). There are 331 communities categorized as *Star* and *Continuous/Grouped* (Fig. 1(A)), including the largest community in this network (whose tooltip is being exhibited in Global View), with 3,886 nodes and 5,035 edges. Fig. 1(D) shows the most summarized version of its node-link diagram. Note that this representation perfectly matches this community's structural categorization (*Star*).

* *The first two authors contributed equally to this work.*

- *C. Linhares, D. Pedro and A. Traina are with Institute of Mathematics and Computer Sciences, University of São Paulo, São Carlos, Brazil. E-mails: {claudiodgl, diogenes.pedro}@usp.br, {agma}@icmc.usp.br.*
- *Luis E. C. Rocha is with Department of Economics & Department of Physics and Astronomy, Ghent University, Belgium. E-mail: luis.rocha@ugent.be.*
- *J. Ponciano and J. Poco are with School of Applied Mathematics, Fundação Getulio Vargas, Rio de Janeiro, Brazil. E-mails: {jean.ponciano, jorge.poco}@fgv.br.*

C Questionnaire for the User Study

The questionnaire was originally written in Brazilian Portuguese, in which all participants were fluent. The questions were translated to English in this document. For the basic and advanced questions, correct or expected answers are marked in bold.

Background and experience

1. Are you aware of any visual difficulties you may have?

2. What area is your education in (e.g., computer science, statistics)?

3. What is your most relevant academic title/function? Choose one option: (i) I'm an undergraduate student; (ii) I'm pursuing my master's degree; (iii) I'm a PhD student/candidate; (iv) I'm a postdoctoral researcher; (v) I'm a professor.

4. What is your degree of prior knowledge in the Information Visualization field? Choose one option: None, Basic, Intermediate, and Advanced knowledge.

5. What is your degree of prior knowledge in the Network Science field? Choose one option: None, Basic, Intermediate, and Advanced knowledge.

6. Briefly explain your experience with the fields mentioned above (Visualization and Network Science).

Algorithm 1 Link length reduction strategy

```
procedure
    if first grid column (timeslice) then
        Insert communities throughout the grid column in the order they appear
    for each remaining grid column X do
        previousColumn ← grid column X-1
        for each "from" community (fromC) in previousColumn do
            for each corresponding "to" community (toC) in X do        ▷ there are two "to" communities in splits
                toPosition ← position in X such that EuclideanDistance(position, fromC) is minimized
                Put toC in toPosition
                Draw link between fromC and toC
        Insert communities not linked so far in the available column positions in the order they appear
        for each "to" community in X that is result of a merge do        ▷ let's minimize merge-related diagonals
            oldFromPosition ← position in previousColumn where its fartest "from" community is located
            newFromPosition ← position in previousColumn where the non-linked community closest to the closest "from" is located
            Switch communities in oldPosition and newPosition
```

Four basic questions (B1 – B4)

Given the Primary School network with default filters, answer:

B1 Considering the "Taxonomy Matrix" with the predefined parameters ("X: Temporal tax." and "Y: Structural tax."), select the two communities categorized as *Clique* and *Continuous and Grouped*. Answer in which positions they are found in the "Global View": (i) (A, 4) and (B, 5); (ii) (F, 12) and (D, 15); (iii) (B, 14) and (C, 16); **(iv) (F, 13) and (D, 14)**; (v) (G, 3) and (B, 4).

B2 In "Global View", select the community (A, 13) and analyze its corresponding "Node-link diagram". This community is composed by the majority of students from school class: (i) 5A; (ii) 2B; **(iii) 5B**; (iv) 1A; (v) 3A.

B3 The community (A, 13) in "Global View" is categorized by the three taxonomies as: **(i) Cont/Disp, Clique and Death**; (ii) Cont/Disp, Star and Birth; (iii) Spor/Group, Circular and Merge; (iv) Spor/Disp, Tree and Clique; (v) Cont/Disp, Tree and Low.

B4 The community (E, 13) in "Global View" is composed of teacher and students from class 4B. In the next timeslice (14), what happens to this community? **(i) Other students join the community, mainly students from class 4A**; (ii) It splits into two other communities; (iii) A group of teachers joins the community; (iv) Students from class 1A join the community; (v) Students from classes 2A and 2B join the community.

Six advanced questions (A1 – A6)

Given the Primary School network with 35 timeslices and communities with at least 15 nodes, answer:

A1 The Primary School network consists of two days, divided by a night period where there is no activity. In addition, there is a lunch break on each day where students from at least three classes meet and interact with each other. Please describe two communities that illustrate expected lunchtime behavior. Answer using the format (row, column), which corresponds to the position of the community in "Global View". **Expected answer: Almost any community between timeslices 5 – 6 and 30 – 32, e.g., (A, 32) and (D, 30). Note that the main challenge here is to find one or two timeslices that is(are) related to the lunch break(s). There are only five correct timeslices among the 35 possibilities.**

A2 Between timeslices 7 and 8 of the "Global View", classes 1A (community (F, 7)) and 5B (community (G, 7)) merge and form a single community (F, 8). You may zoom-in to visualize this behavior better, if necessary. How do you evaluate the statement "It is easy to identify communities that merge"? Choose one option: (i) Strongly disagree; (ii) Disagree; (iii) I don't know; (iv) Agree; (v) Strongly agree.

A3 Based on the Primary School network, do you think it is possible to find other patterns in addition to those mentioned above? If so, please describe one that you consider relevant and mention which part of the visualization helped you find it.

Given the Sexual contacts network with 18 timeslices and communities with at least 5 nodes, answer:

A4 In the "Taxonomy Matrix", select "X: Structural tax." and "Y: Structural tax." and click "Change". Select the 16 communities categorized as Star. A star community is mostly composed of a central node that interacts with the other nodes in the group. Among the 16 communities of this type highlighted in the "Global View", how do you interpret the behavior of the community (F, 9)? Please analyze the node-link diagram and the Temporal Activity Map (TAM) to support your answer. **The expected answer is related to the relationship between one seller and several buyers, with connections spread over time.**

A5 Regarding the community (T, 12), choose the option where all statements are true (if any): (i) It is a Tree. It has only buyers. There are three timestamps with interactions; (ii) It is a Star. It has 3 buyers and 3 sellers. There is a timestamp with five iterations; (iii) It is a Star. It has 1 buyer and 3 sellers. There is at least one interaction at each timestamp; **(iv) It is a Tree. It has 3 buyers and 3 sellers. There is a timestamp where two buyers interact with the same seller;** (v) None of the above.

A6 Based on the Sexual contacts network, do you think it is possible to find other patterns in addition to those mentioned above? If so, please describe one that you consider relevant and mention which part of the visualization helped you find it.

Basic/Advanced questions and intended layouts

Table 1 associates the basic and advanced questions with the visual components they are related to. The basic questions (B1 – B4) and the first advanced question (A1) rely on the use of two visual components to support the participants' answers. In general, Global View was the layout that could be used for all tasks since it represents the overview of the system and leads to other views. Also, since questions A3 and A6 asked the participants to freely explore the system and find new patterns, any layout could potentially be used in this discovery. The second advanced question (A2) requires the analysis of a single component. The last four advanced questions (A3 – A6) depend on the analysis of multiple views.

Questions to compare LargeNetVis with other systems

1. Are you able to do this type of analysis with another system?
2. What other system(s) that allow similar analysis do you know?

Table 1. Basic/Advanced questions and the visual components they are related to.

	B1	B2	B3	B4	A1	A2	A3	A4	A5	A6
Tax mat.	✓		✓				✓	✓	✓	✓
Gl. view	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NL. diag.		✓		✓	✓		✓	✓	✓	✓
TAM							✓	✓	✓	✓

3. What are the advantages of LargeNetVis compared to the system(s) you are familiar with?
4. What are the disadvantages of LargeNetVis compared to the system(s) you are familiar with?

Likert-scale questions (LQ1 – LQ8)

We used two 5-point Likert-scale questionnaires to assess the participants' preferences about LargeNetVis (LQ1 – LQ4) and the provided visual components (LQ5 – LQ8). For each question below, the participants should choose between (i) Strongly disagree; (ii) Disagree; (iii) I don't know; (iv) Agree; (v) Strongly agree.

LQ1 Do you agree with the statement "LargeNetVis offers visual scalability (i.e., it works well for both networks)"?

LQ2 Do you agree with the statement "LargeNetVis is fast (i.e., the provided interactions work in a satisfactory time)"?

LQ3 Do you agree with the statement "LargeNetVis is useful"?

LQ4 Do you agree with the statement "LargeNetVis is intuitive and easy to use"?

LQ5 Do you agree with the statement "The 'Taxonomy Matrix' is useful and helped when analyzing the networks"?

LQ6 Do you agree with the statement "The 'Global View' is useful and helped when analyzing the networks"?

LQ7 Do you agree with the statement "The 'Node-link diagram' is useful and helped when analyzing the networks"?

LQ8 Do you agree with the statement "The 'Temporal Activity Map (TAM)' is useful and helped when analyzing the networks"?

Questions to collect the users' feedback about the system

1. Do you agree with the statement "The three taxonomies and how LargeNetVis uses them contribute to a more efficient network analysis"? Leave a comment.
2. In your opinion, what are the most useful visual aids offered by LargeNetVis? Why?
3. What other visual aids do you think could be useful if incorporated into the LargeNetVis system?
4. Do you have any final comments?

D Primary School - Patterns from question A3

Fig. 2 illustrates some interesting patterns described by participants for the Primary School network (question A3). Fig. 2(I) shows the presence of the biggest network community in the second day of the network. Its existence was pointed out by 30% of participants. This community comprises connections among students from six different grades and three teachers, which was very unusual for this network (see first node-link diagram in Fig. 2). Some assumptions created by the participants were that this pattern represents *"some type of special event"*, or *"an increase of the lunch-break duration"*. Fig. 2(II) shows a community growing between timeslices 27 and 28, which represents an increase in the number of students from the 3rd grade (see the corresponding node-link diagram). At last, two participants indicated that the Sporadic and Grouped temporal categorization guided them to find new patterns on the Global view, which matched the beginning and end of class periods (highlighted in Fig. 2(III)).

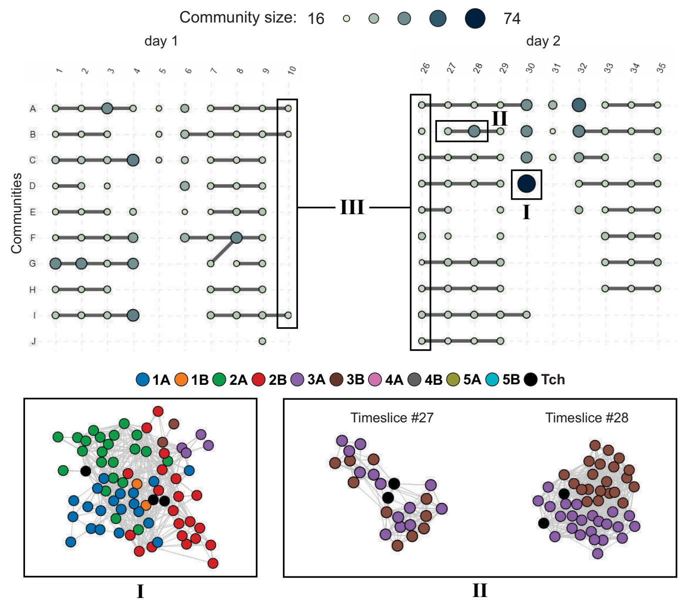

Fig. 2. Primary school: some patterns identified through question A3. (I) Presence of the biggest community in the second day of the network. (II) A community growing between timeslices 27 and 28. (III) Spor/Group temporal categorization matches the beginning and end of class periods. Varying link thickness to map growths and contractions were not available in the version of Global View used in the user study. This feature was implemented based on reviewers' suggestions and feedback from the participants.

E Running Times

Table 2 depicts the average running time of 10 executions for every procedure of the LargeNetVis workflow depicted in Fig. 3(b) of the main paper, i.e., the average running time needed to (1) suggest a range of suitable numbers of timeslices, (2) perform the community detection, (3-5) categorize all communities according to the structural, temporal, and evolution taxonomies, respectively. The experiments were performed on a personal computer with Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz, 16 GB RAM, video card GeForce GTX 1070 8GB, and Windows 10.

Note that the system spent a maximum of 12.66 sec. to run all of these procedures for any network and number of timeslices discussed throughout our usage scenarios and user study. LargeNetVis is still very fast (17.22 sec. on average) when analyzing the Twitter network (discussed in Sec. B). Recall that this network has 50,514 nodes and 108,132 edges. It was decomposed into 100 timeslices and 2,033 communities, numbers many times greater than those from the other networks (Table 2). Every user interaction (zoom, selection, change

Table 2. Average running times (in seconds) for the networks and numbers of timeslices considered in our usage scenario and user study. We also show the average times for the Twitter network mentioned in Sec. B. #C and #T refer to the number of communities detected and number of timeslices considered, respectively. Columns 1-5 refer to the average running time spent to: (1) suggest a range of suitable numbers of timeslices, (2) perform the community detection, (3-5) categorize all communities according to the structural, temporal, and evolution taxonomies, respectively.

Network	#C	#T	1	2	3	4	5	Total (sec)
Primary	81	16	4.33	5.45	0.31	2.55	0.02	12.66
Primary	115	26	4.59	5.73	0.26	1.52	0.03	12.13
Primary	147	35	4.11	5.63	0.23	1.07	0.03	11.07
Sexual	669	18	0,6	5,73	0.43	0.21	0.69	7.66
movie Neg.	154	10	0.7	6.78	0.37	0.24	0.17	8.26
movie Pos.	116	10	0.56	7.45	0.48	0.35	0.14	8.98
Twitter	2,033	100	0.12	12.35	1.46	0.47	2.82	17.22

of colors, etc.) lasts 1 second in the worst case for any network. Both the user feedback and our quantitative analysis demonstrate the well-received response time and computational scalability of LargeNetVis.

References

[1] F. S. Pereira, S. d. Amo, and J. Gama. Detecting events in evolving social networks through node centrality analysis. In Workshop on Large-scale Learning from Data Streams in Evolving Environments of ECML PKDD, pp. 83–93, 2016.