FRCCS 2023

French **R**egional **C**onference on **C**omplex **S**ystems

Le Havre, France

31 May - 02 June

Book of Abstracts

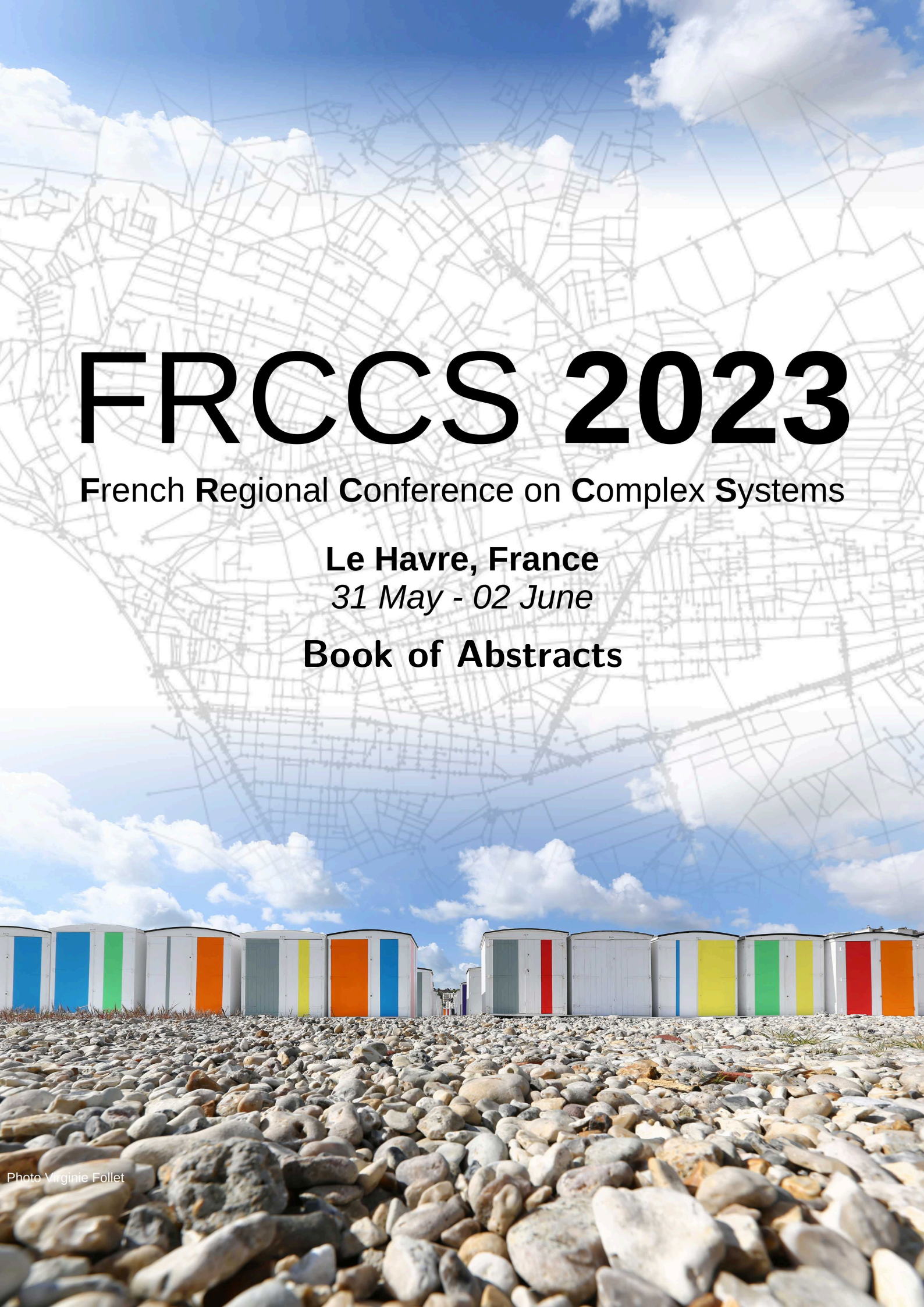


Photo Virginie Follet

Contents

Topics evolution through multilayer networks

Analysing 2M tweets from 2022 Qatar FIFA World Cup

Andrea Russo · Vincenzo Miracula · Antonio Picone

Abstract In this study, we conducted a comprehensive data collection on the 2022 Qatar FIFA World Cup event and used a multilayer network approach to visualize the main topics, while considering their context and meaning relationship. We structured the data into layers that corresponded with the stages of the tournament and utilized Gephi software to generate the multilayer networks. Our visualizations displayed both the relationships between topics and words showing words-context relationship, as well as the dynamics and changes over time by layer, of the most frequently discussed topics.

PACS 05.45.a · 05.65.+b · 01.75.+m

Keywords Multilayer · NLP · Social networks analysis · Football · Data Visualization

1 Introduction

In complexity sciences, given the massive amount of data from the sample under observation, and given the difference in the data collected, a simple layer network cannot suffice to express and show the interactions of many interdependent components in a whole system. Such systems – and the self-organization and emergent phenomena they manifest – lie at the heart of many challenges of global importance for the future of the worldwide knowledge society [1]. For these reasons, in this paper we decided to create a multilayer network [2] with the goal of showing how in a social sample,

Andrea Russo
University of Catania
E-mail: Andrea.Russo@phd.unict.it

Vincenzo Miracula
University of Catania

Antonio Picone
University of Catania

people may change topics depending on various events and time in a given event. The World football cup in Qatar in 2022, was a great opportunity to collect a huge amount of data about about various topics and fans that support national teams. We have collected almost two millions (1.923.283) tweets coming from 2022 Qatar FIFA World Cup (the entire event counts 2.2M tweets) [1], with the #Hashtag "#FIFAWorld-Cup" and "#QATAR2022". In this paper, therefore, we tried to compare how different layers differ from each other, and to get more information about the "context" of the most important topics.

2 Data & Method

We collected the Twitter data thanks to the Twitter API. To create the multilayer network and accurately represent the desired information [3], we began by analyzing the data and then separated the database based on the time of data collection. This enabled us to obtain temporally correct data for each stage of the tournament and visualize it in a 3D dimension [4]. The dates for the tournament stages are listed in Table 1. After separating the various layers, we cleaned the dataset using stop-words. Then to obtain a word network, which could depict the topics but also the context and connected meaning of them over time, we analyzed the data using an algorithm called Bigram. A bigram is a sequence of two adjacent elements from a string of

Stages	Start dates (2022)	Ending dates (2022)
Group stage	20 November	2 December
Round of 16	3 December	6 December
Quarter-finals	9 December	10 December
Semi-finals	13 December	14 December
Final	17 December	18 December

Table 1 Dates used by Stages-Layers

tokens, which are typically letters, syllables, or words. A bigram is an n-gram for n=2. The frequency distribution of every bigram in a string is commonly used for simple statistical analysis of text in many applications, including in computational linguistics, cryptography, speech recognition, and so on. We did not use *TD-IDF* or similar algorithm because we opted for Bigram as the best software for our words-context relation goal. In our code we selected the most used words by users (nodes), and linked them (edges) with the seconds most used words in the same sentence.

We took the most used words that appear for each time-layer, and after several trials[2], we had to choose a limit of 300 words, because it is a good compromise between technical limitations of Gephi[3], and the challenge to posed by a large amount of words with too many nodes and edges, which can make it difficult to understand and not provide useful information such as the texts-contexts.

We used as the layer pillar-words "World", "fifa" and "Team," since they are present in all layers.

[1] https://getdaytrends.com/trend/%23FIFAWorldCup/

[2] https://github.com/AndreaRussoAgid/Multilayer_FRCCS_2023.git

[3] https://answers.launchpad.net/gephi/+question/107399

There are various tools for graphical visualization of multilayers, such as Pymnet[4], but at the graphical level we believe that Gephi is definitely better.

To differentiate layers, with Gephi it is a bit difficult despite having MultiViz [5], in fact even though this tool gives the possibility to make Multilayer starting from some parameters, it is difficult to create a multilayer from the obtained data, because many words are repeated (both at Nodes and Edges level in the various layer). For this reason, we modified the word ID, adding symbols that ransom both the layer and edges of reference between nodes. For the Group stage we did not add any symbols, while for the others (for the purpose of recognition described above) we chose " ^ " for the Round of 16, "*" for the quarters, "†" for the semifinals, and "‡" for the finals.

In conclusion, we initiated the Gephi community algorithm (modularity class) to learn about "word communities" and thus amplify more of the context and meaning of the reference hub and related words.

3 Results & Conclusion

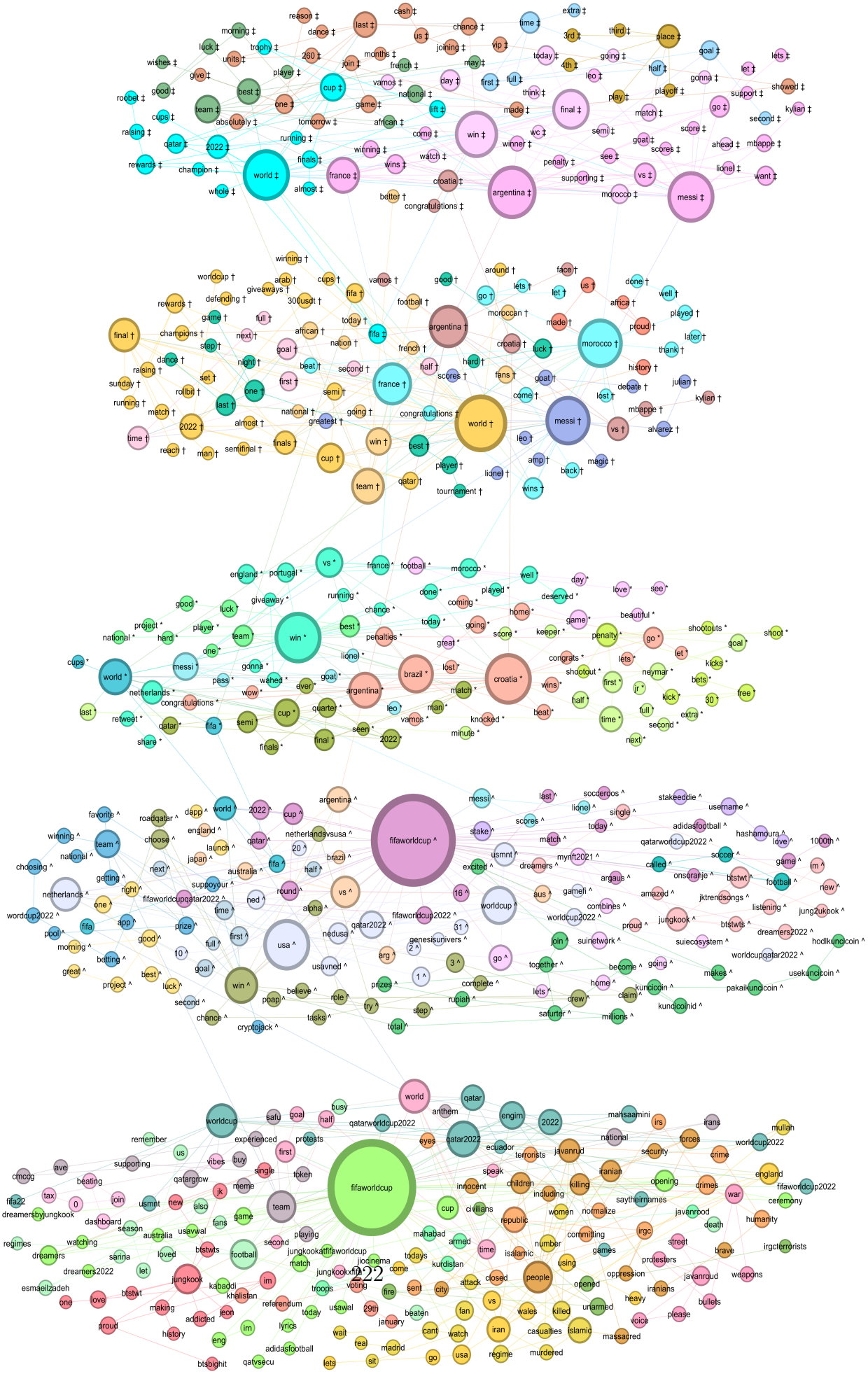
After the cleaning and bigram process, and also the exclusion of nodes and edges that are not part of the main network (gigantic component), we obtained a multilayer network with 858 Nodes and 1041 Edges.

The network depicted in Figure 1 shows that there are changes in the topics discussed across different layers. During the Group stage, word connections that created a discussion context related to crime and social injustices (orange and yellow) and the ceremony's performer, the *BTS* group (in red) were observed. However, as we moved to the subsequent layers (quarters, semifinals, and finals), there was a shift in the focus of the discussion, with less emphasis on the event itself and more on the teams and players. By using Gephi's community algorithm, we identified prominent topics (similar to those of social injustice and BTS) as discussion context. For example, the pride of the Moroccan team (in red †) and the challenge between Messi and Mbappe (in dark pink ‡). We think that the combination of information visualization and multilayer networks could be an effective combination and method for studying the temporal evolution of topics in contexts with multiple levels of interaction.

References

1. S. Boccaletti, G. Bianconi, R. Criado, C.I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, M. Zanin. The structure and dynamics of multilayer networks (2014)
2. M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M.A. Porter, S. Gómez, A. Arenas. Mathematical formulation of multilayer networks (2013). DOI 10.1103/PhysRevX.3.041022. URL https://link.aps.org/doi/10.1103/PhysRevX.3.041022
3. M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Journal of complex networks **2**(3), 203 (2014)
4. F. Mcgee, M. Ghoniem, G. Melançon, B. Otjacques, B. Pinaud, in *Computer Graphics Forum*, vol. 38 (Wiley Online Library, 2019), vol. 38, pp. 125–149
5. J.P.C. S., A. Chatterjee, G. M., A. Mukherjee. Multiviz: A gephi plugin for scalable visualization of multi-layer networks (2022). DOI 10.48550/ARXIV.2209.03149. URL https://arxiv.org/abs/2209.03149

[4] http://www.mkivela.com/pymnet/


world ‡
qatar ‡
2022 ‡
cup ‡
team ‡
win ‡
final ‡
france ‡
argentina ‡
messi ‡
morocco ‡
croatia ‡
world †
qatar †
2022 †
cup †
team †
win †
final †
france †
argentina †
messi †
morocco †
croatia †
world *
qatar *
2022 *
cup *
team *
win *
final *
france *
argentina *
messi *
morocco *
croatia *
fifaworldcup ^
qatar ^
2022 ^
cup ^
team ^
win ^
usa ^
fifaworldcup
qatar
2022
cup
team
win
usa
world
messi
argentina
france
croatia
morocco