Hello I'm a title

Isabela Constantin, Adélie Garin, Celia Hacker, Michael Spieler

I. Introduction

Using the wikipedia data that is now open source, we would like to understand how the number of views on some pages is dependent on the type of event concerned. We will consider two types of events :

1) Expected events, e.g. elections, concerts, political events,...
2) Unexpected events e.g. death of someone famous, coup d'état, natural disaster...

Choosing several specific events of each type, we will consider the graph built with pages concerned as nodes and a two pages are linked if there is a web link from one to the other. We will use both the directed graph, in which one page which is linked to the other leads to only one direction for the concerning edge, and the undirected graph, for which there exists an edge if and only if there is one page linked to the other (we do not ask for the reciprocal page to be also linked to the other). Our question is: Does the number of views on

the pages propagates in the pages in different ways depending on the type of events? If yes, how do they behave? We will

consider the difference between the number of views of the day before the event and the day of the event. To analyse our

data and try to answer our questions, we will first start by acquire the data, by building several graphs for each type of events, then explore them and try to get general properties of the graphs, building a general pipeline for our analysis. We then exploit the data, trying to test our hypothesis and finally draw conclusions.

II. Data Acquisition

We constructed the graphs by selecting the wikipedia article concerning each event. We then grew the graphs around them by selecting the pages the event was linking to, and the pages linked from those. To reduce the amount of data we had, we randomly subsampled by giving a higher probability to stay in the graph to the nodes that are higher in the page, meaning they are stated early, so they are more likely to be important links. The resulting graphs are directed, unweighted and connected. In addition, we found the number of views per page, that we computed for blablabla.

III. Data Exploration

We start by analysing some basic properties of each graphs. The results are stated in the following table.

TABLE I
BASIC PROPERTIES OF THE GRAPHS

Graphs	Properties			
	Number of nodes	*Number of edges*	*Diameter*	*Average Clustering Co*
copy	More table copy[a]			

[a]Sample of a Table footnote.

IV. Conclusion

References

[1] [2] [3]

References

[1] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
[2] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.
[3] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.