LABORATOIRE
BORDELAIS
DE RECHERCHE
EN INFORMATIQUE
LaBRI

université
de BORDEAUX

Document annexe au rapport sur la thèse de Alexis Pister intitulée:

**" Analyse Visuelle de Réseaux Sociaux Historiques:
Traçabilité, Exploration et Analyse "**

Ce document reprend un ensemble de remarques notées au fil de ma lecture. Toutes ne demandent pas à être prises en compte.

Je n'ai pas relevé les coquilles de manière exhaustive, et ai pu baissé la garde à mesure que j'avançais dans ma lecture …

Intro and related work section

Page 1, SNA used in geography misses a reference. How about using:

> Rozenblat, C. and G. Melançon, Eds. (2013). Methods for Multilevel Analysis and Visualisation of Geographical Networks. Methodos Series.

Page 5, line 17, « Most visual software for SNA such as […] » -> « Most used visual software for SNA such as […] »

Page 5, au risque de paraître chauvin, je m'attendais à trouver Tulip parmi la liste des « Most visual software for SNA such as […] » (qui est cité plus loin – au bas de l page 15)

> Auber, D., et al. (2018). Tulip 5. Encyclopedia of Social Network Analysis and Mining. R. Alhajj and J. Rokne. New York, NY, Springer New York: 3185-3212.

Page 26, ligne -1, Strictly speaking, you don't say what the set T is and what data a timestamp corresponds to.

Page 34, you mention your publication at VIS4DH is available through IEEExplore. You should provide a DOI in the bibliography (I could not locate the document on IEEExplore, I had to look at the version stored on HAL).

Chapter 4 CombiNet

Page 37, section 3.3. Discussing the several pitfalls that may occur in the acquisition/digitization/annotation/analysis process, we may expect the « volume » of documents (the amount of exploitable data they contain) to translate into the complexity or at least size of the considered network. Shouldn't this be taken into account as it certainly has some impact on the « severity » of the potential pitfalls?

Page 53, line -24, citing Lemercier et al. While it is true that the « spaghetti monster » by itself may be worthless, techniques were designed to filter graphs (based on their topology). One reference I believe you ought to mention is the work by Bobo Nick et al.:

> Nick, B., et al. (2013). Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks. Advances in Social Network Analysis and Mining (ASONAM), Niagara Falls, Canada.

Page 54, line -1, when mentioning that the number of graphlets grows exponentially, it might be interesting ti refer the curious reader to the Slaone integer sequence from where he could find (numerous) specific references. (although graph lets seem to only consider connected graphs).

Page 58, line -7, so it seems users need to go back to the original documents (to find specific information « What price does the emigrant have to pay, given their socio-economic profiles? »).

351, cours de la Libération
F-33405 Talence Cedex
Tél.: +33(0)5 4000 69 00
Fax: +33(0)5 4000 66 69

LABORATOIRE
BORDELAIS
DE RECHERCHE
EN INFORMATIQUE
LaBRI

université
de BORDEAUX

Chapter 5 Prior Knowledge Clustering

Page 92, line 7, the reference to « Bipartite Modularity Optimisation [] » is missing.

Page 92, line 8 you provide a reference when mentioning one-mode-projection. I didn't take time to review this reference, but in my opinion projection onto one-mode network (from a 2-mode network) is folklore. I don't see what is specific to the work by Zhou et al. that makes it relevant to be cited here.

Page 97, in Fig. 5.5, colors pink (resp. blue) is assigned to cluster A (reps. B). The legend however seem to show the inverse mapping (A -> blue, B -> pink). Or did I miss something?

Page 100, line -1, a reference to the « network about the trades of Marie Boucher [] » is missing.

Pages 105 et ss. (section 5.5). It would have been interesting, and just as valuable, to develop a situation where prior knowledge is contradicted by a majority of algorithms (no consensus at all) which could indicate that whether no algorithm is able to properly identify a clustering structure in the data, or that prior knowledge is biased and unverified. A similar situation could be that consensus is reached by a single or a minority of algorithms which may question the robustness of the process.

Conclusion

Page 111, line -20. The quote from Headland et al. « document nodes can be considered as Emics and person nodes Etics concepts » is put there without any further development, leaving the curious but ignorant reader with a question mark. The discussion is worth explaining what is meant and how it can be taken into account in future work.

Page 113, line 1, you mention the fact that CombiNet « does not allow complex network transformations (such as creating simple unipartite networks from projections) ». I insist that you expand this remark by referring to Borgatti (and eventually to Renoust's EuroVis 2015 paper).

Coquilles (typos)

Page 2, « […] some are which still relevant. » -> some of which are still relevant (?)

Page 5, ligne 10, « As freeman writes […] » -> As Freeman writes […]

Page 7, ligne -7, « (dense, parse, connected, etc.) » -> (dense, sparse, connected, etc.) (?)

Page 17, ligne -2, « […] farmers, workers, property owners etcinstead of solely […] » -> farmers, workers, property owners etc. instead of solely (?)

Page 18, ligne 13, « Despite the controversy, These kinds of approaches […] » -> Despite the controversy, these kinds of approaches

Page 17: the sentence on line 1 sounds awkward (« study of documents to study the past » :-)
« Nevertheless, History can be characterized as the collection and study of historical documents to study and describe the past. ».

Page 23, ligne 11, « As defining what an important node is ambiguous » que l'on a envie de lire comme « As defining what an important node is is ambiguous » (?), avec une façon de se tirer d'embarras « Because the definition of an important node is ambiguous » :-)

Page 34, line 9-10: « These principles emerged from joint experiences as historians and computer scientists while collaborating on multiple projects » -> These principles emerged from joint experiences as historians and computer scientists were collaborating on multiple projects

Page 39, line 4-6, the sentence « However, before mid 20th century, most historical primary sources are stored in archives in paper format and need human work to be digitized. » appears twice in a row.

351, cours de la Libération
F-33405 Talence Cedex
Tél.: +33(0)5 4000 69 00
Fax: +33(0)5 4000 66 69

Page 90, line -6, I am unsure whether the word « appartenance » is good english (American or UK?) An alternative may be: « Similarly, users are not required to assign every node to a cluster, as social scientists often do not have a strong opinion about whether individuals belong to a group. »

Page 113, line -13, I am no native english reader, but I feel the end of the following sentence is ill-formed: « which is the layout PK-Clustering is based on. » (Did you mean « on which the layout of PK-Clustering is based » ?)

351, cours de la Libération
F-33405 Talence Cedex
Tél.: +33(0)5 4000 69 00
Fax: +33(0)5 4000 66 69