

SA/DA/MA: The World of Music

Have you ever thought about how the world of music looks like? For instance, where would you place Rolling Stones? Or, how close do you think the Beatles are to Mozart or Nirvana? Are they neighbors or are they far apart from each other, like Che Guevara and Roger Wattenhofer, for example?

If you have never thought about it, don't worry - other people have! There are online databases, like *allmusic.com*¹, for example, that contain various dependencies collected by many researchers. Such information is the base of several existing applications, such as music suggestion systems or play-list generators. Many new applications are expected to appear in the near future, once all these relations are better understood.

One way to better understand the structure of the World of Music is to think about these dependencies as a graph, where artists (or songs) are nodes and edges represent direct relations. If such a graph is at your disposal, well known methods, such as embeddings, shortest path calculations, or the HITS algorithm, could be used to reveal interesting information. Unfortunately, the data of *allmusic.com* is not freely available, which makes it difficult to obtain such a graph.

In this thesis, your task is to find ways to construct the graph of the World of Music, such that it is accessible for further research. Different approaches might lead to the goal: Maybe we could *crawl* the website of *allmusic.com* (scan the hyperlink structure by a robot). However, the administrators of this website don't really like this approach and would block our IP address after a few thousand consecutive requests. Thus we would have to do it cleverly – in a peer to peer manner maybe? Another option might be to extract the data from Wikipedia², which contains an astonishing amount of information about music. Yet another approach might be to set up the graph based on usage statistics, possibly taken from the *Last.fm*³ site (will they allow us to crawl their data?). And, of course, you might have yet another brilliant idea!

Interested? Please contact us for further details!

Kind of work: Mostly practical

Contact:

1. Kuhn Michael: kuhmi@tik.ee.ethz.ch, ETZ G61.4, phone 044 632 77 30
2. Olga Goussevskaia: golga@tik.ee.ethz.ch, ETZ G61.4, phone 044 632 70 05
3. Roger Wattenhofer: wattenhofer@tik.ee.ethz.ch, ETZ G63, phone 044 632 63 12

¹ www.allmusic.com

² www.wikipedia.org

³ www.last.fm



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Prof. Roger Wattenhofer
<http://www.dcg.ethz.ch>

SA/DA/MA: Similarity Measures in Small World Graphs

In 1967 the sociologist Stanley Milgram conducted a famous experiment: He let randomly selected people from the US west coast send a letter to other, also randomly selected people on the east coast. The participants were thereby only allowed to send the letters to friends who in turn could send them to their friends and so on. The experiment revealed that a chain of friends from the US west to the US east coast consists of astonishingly few links. This observation led to the famous expressions “six degrees of separation” and “small world” that triggered a huge amount of research activities.

Surprisingly, not only social graphs (i.e. the graph resulting from relationships between people), but many other graphs from everyday life, such as the WWW-graph, the graph of publications, or the graph resulting from the link structure of Wikipedia have very similar characteristics. An important property such graphs have in common is that they exhibit an extensive clustering among some nodes and at the same time contain edges that span over many clusters and thereby produce a small overall diameter of the graph. For many applications it would be interesting to have some understanding of the relatedness of nodes within such a graph. E.g., imagine a graph representing the dependencies between songs. Having a good measure for the relatedness of songs within such a graph could lead to more sophisticated file-sharing tools or play-list generators, or to even more fancy applications such as predictors of trends within music.

Unfortunately, a simple shortest path measure does not immediately lead to the desired result. In this thesis you are therefore asked to compare different methods that could define such a measure. Could cleverly assigned edge weights improve the shortest path approach considerably? Could ideas used to train neural networks help to train edge weights? Is the network flow a good measure? How can we quantify the quality of a measure in order to compare the different methods? Depending on your interests and abilities, you can tackle the topic from a mostly experimental (i.e. practical) or a rather theoretical side.

Interested? Please contact us for more details!

Contacts:

1. Kuhn Michael: kuhnmi@tik.ee.ethz.ch, ETZ G61.4, phone 044 632 77 30
2. Yvonne Anne Oswald: oswald@tik.ee.ethz.ch, ETZ G61.2, phone 044 632 59 86
3. Roger Wattenhofer: wattenhofer@tik.ee.ethz.ch, ETZ G63, phone 044 632 63 12