

MUSIC RECOMMENDATION MAPPING AND INTERFACE BASED ON STRUCTURAL NETWORK ENTROPY

Justin Donaldson

Indiana University School of Informatics
Human Computer Interaction/Design

Ian Knopke

Indiana University School of Informatics
Music Informatics

ABSTRACT

Recommendation systems generally produce the results of their output to their users in the form of an ordinal list. In the interest of simplicity, these lists are often obscure, abstract, or omit many relevant metrics pertaining to the measured strength of the recommendations or the relationships the recommended items share with each other. This information is often useful for coming to a better understanding of the nature of how the items are structured according to the recommendation data. This paper describes the ZMDS algorithm, a novel way of analyzing the fundamental network structure of recommendation results. Furthermore, it also describes a dynamic plot interaction method as a recommendation browsing utility. A novel "Recommendation Map" web application implements both the ZMDS algorithm and the plot interface and are offered as an example of both components working together.

1 INTRODUCTION

Item based recommendation systems such as the music recommendation services offered by MyStrands¹ are based on networks of associations between items. These associations are formed from aggregate observations or records of sets that occur between the items. In the context of the MyStrands music recommendation network, these sets are playlists. A network can be constructed from this data where each item is a node and the connection weights between the items are the number of times that the two items appear in any set. Recommendation for music in this context involves providing one or more songs as "input" into the system. The system resolves the songs against its list of known songs, and then retrieves a set of recommended songs that have strong associations for the provided songs.

For the purposes of this project, a dataset of 1.6 million songs from the MyStrands recommendation database was used. The data was constructed from hundreds of thousands of playlists, which in turn formed millions of links between the individual songs. Recommendation results were retrieved through a modified network neighborhood

¹ <http://www.mystrands.com>

extraction routine. This process resolves the input songs in the network, and then retrieves each of the direct neighbors of these songs.

2 MUSIC NETWORK VISUALIZATIONS

Since the recommendation data is a network, it is possible to visualize the prominent structure of the network in two or three dimensions using conventional network visualization techniques. Many recent projects have attempted to visualize music relationships in a low dimensional space according to association or content based similarity metrics [2, 1]. Visualizations such as these are extremely valuable for understanding the various relationships in music.

3 ZMDS AND STRUCTURAL NETWORK ENTROPY

ZMDS is a novel modification of the standard Euclidean distance based multidimensional scaling method. It involves the following steps: (1) Construction of a modified association matrix from the recommendation result, (2) row-wise z-score normalization of node edge weights (The Z in ZMDS is taken from this step.), (3) row-wise Euclidean distance calculation for dissimilarity matrix, and (4) classic multidimensional scaling method on dissimilarity matrix for required number of dimensions.

According to this method, in step 1 a matrix of associations from a recommendation result is constructed. This matrix is similar to a Laplacian matrix, where each non-diagonal entry is the association weight between two of the songs in the recommendation result, and each diagonal entry is the global edge weight for the song.

Each song's connection weights in the matrix are modified by a function of the songs total participation in the network in step 2. In order to view the resulting structure in Euclidean space, it is necessary to convert the asymmetric weighted matrix into a symmetric version. This can be done by calculating row-wise Euclidean distances and constructing a symmetric dissimilarity matrix. Applying a multidimensional reduction algorithm on the dissimilarity matrix will provide a low dimensional representation fit for visualization. A z-score normalization method is preferred which uses variance (unit standard deviation) as



Figure 1. Artist, genre, and popularity clustering in ZMDS graph

the basis for edge weighting:

$$w_{i,j} = \frac{k_{i,j} - \bar{k}}{\sigma} \quad (1)$$

With $w_{i,j}$ being the node weight between nodes i and j , $k_{i,j}$, being the co-occurrence weight, and σ being the standard deviation for row k .

By encoding the size of the node by its global edge degree, it is clear to see how the large hub-pop songs are marginalized on the left side of the distribution. The right side of the distribution describes the two tail structures for the recommendation set. In the context of the recommendation set retrieved, these tails correspond to songs by the artist Bruce Springsteen and Tori Amos (See Figure 1). These artists have long histories and loyal fan bases. As a result, they have a catalog of songs that share more intra-catalog edges than extra-catalog edges. However, the associations do not form a cluster in the traditional sense. Instead, the tail structures can be thought of as a “gradient” cluster feature. Nodes along the ends of the tails will include regions of songs that are less widely popular for a given artist. However, while most of these songs are only associated with other Springsteen and Amos songs, there are songs that act as bridges or gateways into the rest of the music network. Logically, these songs are the most widely popular songs for both artists. Near the center of the distribution where these tails start is the “low entropy hub” which ties together the tails into the larger high entropy fan region on the left. It is important to note that one or more of these features may be missing from a ZMDS distribution. If there is no significant recommendation set structure, there will be no tails. Likewise, low entropy hubs may be missing as well.

4 INTERACTIVE VISUALIZATION

A node repulsion technique was used to handle the node occlusion that occurs in such representations. This technique is very similar to a class of techniques collectively known as liquid browsing [3]. The motivation for the interactive implementation is to allow a user to recognize



Figure 2. Recommendation mapping applet

and investigate occlusion as it occurs in the low dimensional representation, as well as to retrieve non-dimensional information about the node, such as artist or track title. The visual interface was created in Macromedia Flash as a lightweight web application, suitable for visualizing and investigating music recommendation sets of around two hundred elements. The actual ZMDS process is performed using the PDL matrix data language², and passed to the Flash applet as an XML data set (See Figure 2). The overall effect of this behavior is easy to perceive in the context of the web interface itself, which is available online³.

5 REFERENCES

- [1] F. J. Igo, M. Brand, K. Wittenburg, D. Wong, and S. Azuma. Multidimensional visualization for collaborative filtering recommender systems, 2002.
- [2] E. Pampalk. Islands of music: Analysis, organization, and visualization of music archives. *Journal of the Austrian Soc. for Artificial Intelligence*, 22:20–23, 2003.
- [3] C. Waldeck, D. Balfanz, C. G. Center, and G. ZGDV. Mobile liquid 2d scatter space (ml2dss). *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 494–498, 2004.

² PDL: The Perl Data Language: <http://pdl.perl.org/>

³ Recommendation Mapping Applet: <http://labs.mystrands.com/cgi-bin/recmap.cgi>