# Detecting Influencers in the YouTube Market

Emily Muller (emily@aims.ac.za)

August 17, 2017

This document is intended for reading alongside the jupyter notebook 2017-08-05-network-from-scrape.

## 1 Network

The sub-setted network of influential YouTube users will be modelled as a weighted directed network, where nodes are the users (or sellers) and an edge from user $U_i$ to user $U_j$ is a measure of interaction. The following notation will be used:

| Notation | Description | Domain |
|---|---|---|
| $U_i$ | users/sellers in the network | $i = 1, ..., n.$ |
| $I_{jk}$ | items in the market (videos) from user $U_j$ | $k = 1, ..., \|I_j\|.$ |
| $e_i e_{jk}$ | $\begin{cases} 1 & \text{there exists interaction (comment) from user i to item k of user j} \\ 0 & \text{there is no such interaction} \end{cases}$ | {0,1} |

Consider first the set of users and videos. These can be modelled as a bipartite graph, with $U_i$ on the left, $I_{ik}$ on the right and edges, $e_i e_{jk}$, directed from left to right.

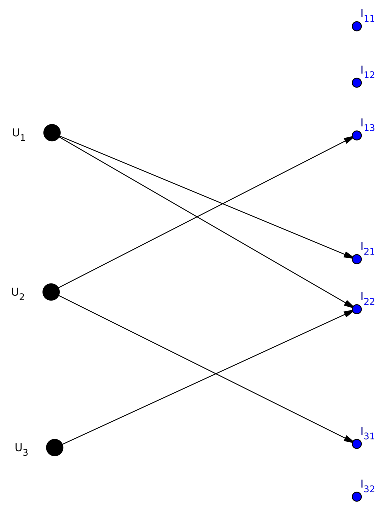


Figure 1: Example of bipartite graph representing network

Figure 1 shows an example of a network consisting of 3 users. Note that an interaction from $U_i$ to $I_{jk}$ is only counted once for each comment. The aim is to derive from the bipartite graph, a weighted directed graph depicting interaction activity between users. It is necessary to consider both, the number of possible interactions from user $U_i$ to $\{I_j\}$ and the activity of user $U_i$. A method similar to that used in calculating recommender power for collaborative filtering recommender algorithms is used [?]. The interaction from $U_i$ to $U_j$ is defined as,

$$I(U_i, U_j) = \frac{\text{\# of edges from } U_i \text{ to } I_{jk} \text{ for all } k}{\text{total \# edges sent from } U_i} \cdot \frac{\text{\# of edges from } U_i \text{ to } I_{jk} \text{ for all } k}{\text{total \# items } I_{jk} \text{ for all k}} \tag{1.1}$$

$$I(U_i, U_j) = \frac{\sum_{k=1}^{|I_j|} e_i e_{jk}}{\sum_{j=1}^{n} \sum_{k=1}^{|I_j|} e_i e_{jk}} \cdot \frac{\sum_{k=1}^{|I_j|} e_i e_{jk}}{|I_j|}. \tag{1.2}$$

Therefore, the interaction matrix for Figure 1 is,

$$I = \begin{bmatrix} 0 & 1 & 0 \\ 1/6 & 0 & 1/4 \\ 0 & 1/2 & 0 \end{bmatrix}. \tag{1.3}$$