Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 655 – 659

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014-Philadelphia, PA

# Organizing Patterns and Evolution of Indian Movie Industry

Srinivasan Radhakrishnan[a*], Rohit Jacob[b], Arjun Duvvuru[c], Sagar Kamrthi[d]

[a]Academic Research Associate, Symbiosis Institute of Management Studies, a constituent of Symbiosis International University
[a]Research Assistant, Symbiosis Institute of Management Studies, a constituent of Symbiosis International University
[c]Quality Engineer at JDA Software Pvt Ltd, India
[d]Professor, Department of Mechanical and Industrial Engineering, Northeastern University

**Abstract**

In this study we focus on basic structural and temporal characteristics of a network formed by linking movie actors in accordance with their co-occurrences with other fellow movie actors. In an actor network, the actors are represented as nodes and a link is formed between a pair of actor if they appear in the same movie. Each link is assigned a weight, representing the number of co-occurrences of the pair in different movies. A statistical analysis of the structure of networks reveals the organizing pattern of movie actors and temporal analysis of networks shows the evolution of the movie actors and industry on a whole. In this study we compare the national and regional movie industry in India along selected weighted network measures. The data set contains movies (from 1980 to 2013) encompassing three movie industries, namely Bollywood (national), Tamil and Malayalam (regional). We observe strong agreement between the three movie industries for degree distribution, strength distribution, average weight as a function of degree, and weighted clustering coefficient as a function of degree. The aforementioned observance does not deviate in temporal scale. However we observe a key difference in affinity measure between national and regional movie industries suggesting more acceptance of new actors in case of regional movie industry.


Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

*Keywords:* Complex Networks; Movie Actor Networks; Weighted Networks

## 1. Introduction

Movie and society form a closed influence loop where both impart change (positive & negative) on to each other. The contents of the movie may impact the awareness pertaining to a particular subject or provide pure entertainment. Emergence of digital medium and strong distribution channels have ensured wide and efficient spread of country specific movies across the globe. A regional movie focusing on regional issues is no longer constrained by geographical boundaries. Although the right recipe for witnessing the effects of global movie synergies are being experimented upon, the basic evaluation parameters of a movie's performance remains the same. The performance of a movie is mainly measured by earnings (either global or worldwide) and recognition (in form of awards) with more emphasis on the former. The earnings based evaluation reveals high coupling of the movie's success and its lead actors. This method fails to capture the contribution of the non-lead actors and thereby makes it difficult to analyse the holistic evolution of the movie industry. On the other hand, the recognition based evaluation gives some visibility to the non-lead actors but not sufficient enough to surpass the lead actor contributions. Scenarios exploiting both


Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2014.09.070

earnings and recognition based evaluation end up in favour of lead actors leaving behind a murky trail of non-lead actors and inability to grasp holistic evolution of the movie industry.

We develop a complex network based analysis method which will enable us to track the evolution of actors and highlight the global behaviour of the movie industry. In this study we focus on basic structural and temporal characteristics of network formed by linking actors in accordance with their co-occurrences with other fellow actors. In actor network (see Figure 1), the actors are represented as nodes and a link is formed between a pair of actor if they appear in the same movie. Each link is assigned a weight, representing the number of co-occurrences of the pair in different movies.

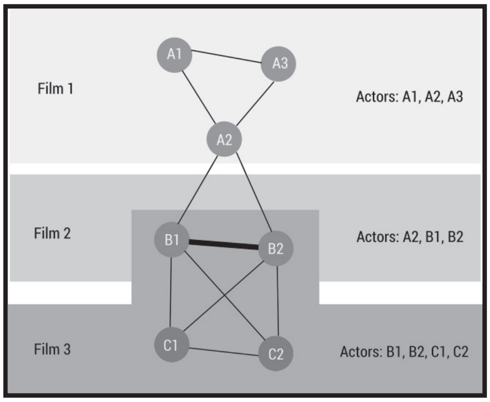


**Figure 1:** Formation of actor network

Co occurrence networks are most suited to reveal the evolution of a system which has finite set of entities with non-zero probability of establishing a link between them. Duvuru.et.al[2], extensively used co occurrence networks to uncover emerging trends in academic research. The weighted nature of co occurrence networks calls for network measures specific to weighted networks. Applying network measures designed for unweighted to weighted network may not yield better results. A set of network measures was designed by Barrat et al[1]. for weighted networks which showed superior representation of the network's structural characteristics. We now discuss several relevant network measures used for analysis of the weighted network formed using aforementioned technique.

**Degree Distribution:** Degree of a node is the number of links incident on the node. It is a measure of centrality or relative importance of a node in the network. Degree distribution is a frequency plot of the nodes' degrees in a network. Degree distribution provides an overall picture of the structure of a network. Most real world networks follow a power-law distribution, signifying the presence of a considerable number of high degree nodes or hubs which form the backbone of a system. Degree distribution of a network will determine if the structure and behaviour is random or scale-free.

**Strength Distribution:** The degree of a node may intuitively hint at the importance of the same. However it may not necessarily be a better measure in weighted networks[1]. A weighted network is often described by a weighted adjacent matrix $w_{ij}$, which represents the weight on the link between node *i* and *j*, with $i; j = 1; : : : ; N$, where *N* is the size of the network. Here, only the undirected network with symmetric weights $w_{ij} = w_{ji}$ is considered. The definition of degree can be extended to strength naturally, as

$$Si = \sum_{j \in \Gamma(i)} w_{ij}$$

where the sum runs over the set $\Gamma(i)$ of neighbours of node *i*. The strength of a node integrates the information about its degree and the weights along its links, thus can characterize the importance of the node more accurately than its degree.

**Average Weighted Nearest Neighbour's Degree as a Function of Degree:** Weighted average nearest-neighbors degree, defined as given below[1]

$$k^{w}{}_{nn,i} = \frac{1}{S_i} \sum_{j=1}^{N} a_{ij} w_{ij} k_{j}$$

A local weighted average of the nearest-neighbour degree is considered according to the normalized weight of the connecting edges, $w_{ij} / S_i$. The $k^w{}_{nn,i}$ measures the effective affinity to connect with high- or low-degree neighbours according to the magnitude of the actual interactions. As well, the behaviour of the function $k^w{}_{nn}(k)$ highlights weighted assortative or disassortative properties[1].

**Average Weight as a Function of End Point Degree:** This measure is defined as follows

$$< w_{ij} > \sim \left(k_i k_j\right)$$

This measure allows one to observe co-occurrence of link between two nodes as the degree changes. In this work it will enable us to observe the how frequently high degree actors act with each other.

**Weighted Clustering Coefficient as a Function of Degree:** This measure is defined as[1]

$$C_i^w = \frac{1}{S_i\left(k_i - 1\right)} \sum_{j,h} \frac{\left(W_{ij} + W_{ih}\right)}{2} a_{ij} a_{ih} a_{jh}$$

Weighted coefficient is a measure of the local cohesiveness that takes into account the importance of the clustered structure on the basis of the amount of traffic or interaction intensity actually found on the local triplets. $C_i^w$ counts for each triplet formed in the neighbourhood of the vertex *i* the weight of the two participating edges of the vertex *i*. This records the number of closed triplets in the neighbourhood of a vertex and their total relative weight with respect to the strength of the vertex. This is a measure of how cohesive (well connected to each other) a group of nodes are or how well connected a node is to its neighbours[1].

## 2. Data

Data of all the actors who acted in movies from 1980 as on date was collected for Bollywood (national), Tamil and Malayalam industry (regional). In all, data for 15,326 movies were collected. The data was clustered in set of five years (1980-1984, 1985-1989, 1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2013). In all seven groups were formed based on temporal classification for data pertaining to each of Bollywood, Tamil and Malayalam industry.

## 3. Analysis

The data collected was analysed along selected network measures. We observe that the Bollywood actor network is topologically similar to the scientist collaboration network[1] (highly assortative and lack of clustering of hubs). The Tamil and Malayalam actor network is similar to the knowledge network[3] (highly disassortative and lack of clustering of hubs).

**Degree Distribution:** Figure 2A shows the degree distribution of the movie actor network for all the three regional segments. We find power-law degree distribution across Bollywood, Tamil and Malayalam industries for all the five year group periods from 1980 to 2013. It points to the scale-free nature of the movie actor network which indicates presence of few highly connected nodes (hubs) and many nodes with few connections. Moreover the tendency to deviate from power law across the temporal frame is absent. It is important to note that the hubs do not exclusively represent popular lead actors. In this case we observe that the hubs are a mix of popular lead actors and frequent supporting actors.

**Strength Distribution:** We observe similar trends in strength distribution (see Figure 2B). We find power-law strength distribution across Bollywood, Tamil and Malayalam industries for all the five year group periods from 1980 to 2013. It points to the scale free nature of the movie actor network which indicates presence of few nodes with high strength and presence of lots of nodes with low strength values.

**Average Weighted nearest Neighbour's Degree as a Function of Degree:** Figure 2C shows the change in average weighted nearest neighbour's degree as a function of degree. We observe negligible deviation that across temporal scales within each industry. However Bollywood industry produces more sharp increase in weighted nearest neighbour degree as the degree increases. For Tamil industry the increase in the value retains for lower degree (degrees less than 5). For Malayalam industry the value shows negligible variation across degrees. This measure reflects the *affinity* of a node to connect with similar node. In other words it reflects assortative behaviour. In case of Bollywood industry the high degree nodes tend to connect to other high degree nodes. However strong disassortative behaviour is noticed in case of Tamil and Malayalam movie industry. This indicates the ability of Tamil and Malayalam industry to absorb new actors (and actors who have worked with few other actors) in a common movie with established actor who has acted in more number of movies and with more artists.

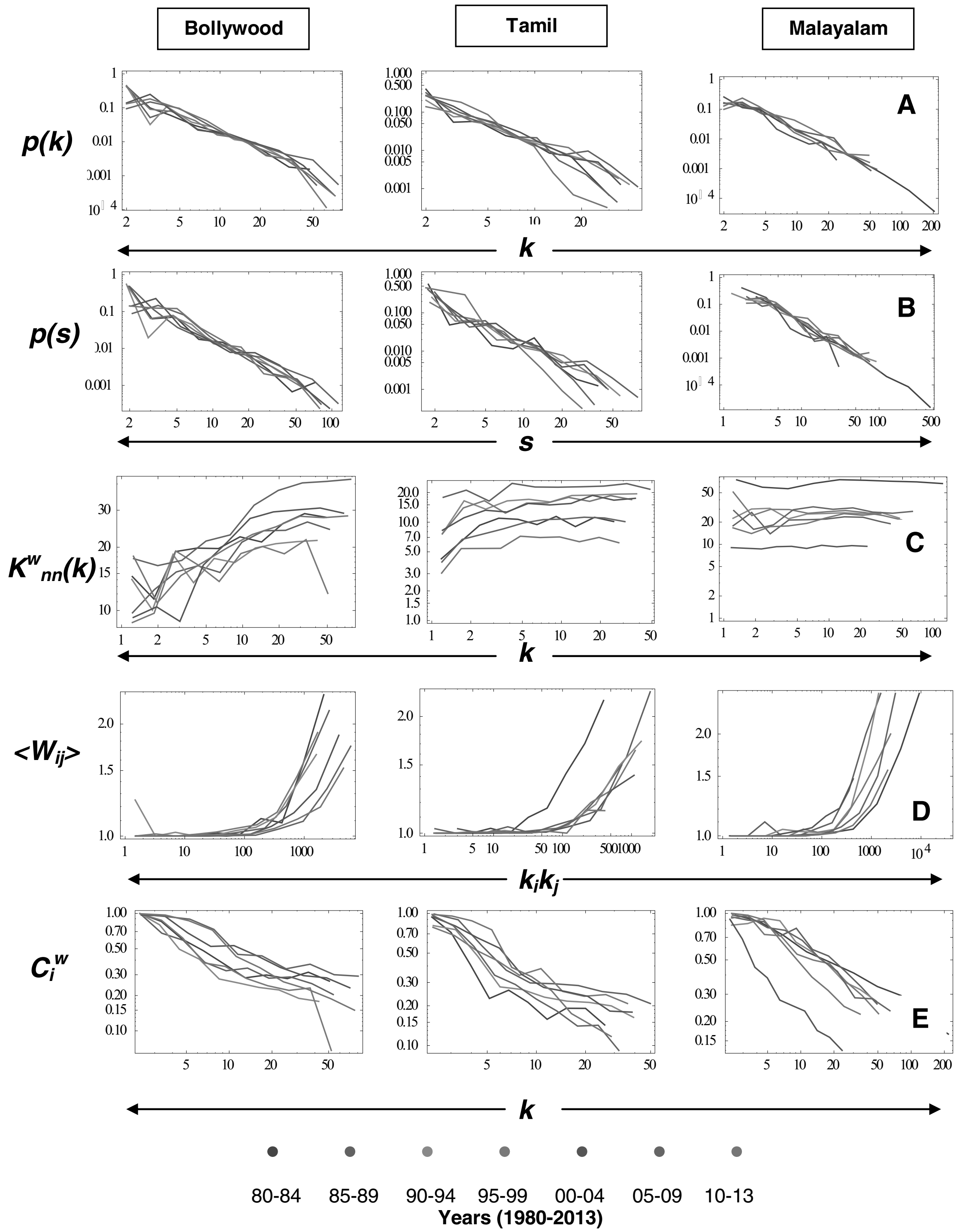


**Figure 2:** Comparison of selected network measures across preset temporal frames starting from 1980 to 2013. Colours indicate the temporal window. The following network measures are compared between actor network from Bollywood, Tamil and Malayalam movie industry: A) Degree distribution, B) Strength distribution, C) Average weighted nearest neighbour's degree as

a function of degree, D) Average weight as a function of end point degree, E) Weighted clustering coefficient as a function of degree.

**Average Weight as a Function of End Point Degree:** Figure 2D shows how average weight changes in accordance with the end point degree. The average weight increases in accordance with increase in end point degree. This phenomenon is observed across all the three movie industries across all temporal frames. All the three industries exhibit an exponential increase in values of average weight ($<w_{ij}>$) for values of $k_i k_j > 100$. It indicates that the tendency to act in multiple movies together increases between high degree actors (the actors who have acted with more number of actors tend to work with other actors who have worked with more actors).

**Weighted Clustering Coefficient as a Function of Degree:** Figure 2E shows the weighted clustering coefficient as a function of degree. The observations on decay in values are similar across all the three industries and across all temporal frames. This reflects less clustering in hubs than lower degree neighbours. Actors with few connections prefer to work in close knit communities. Actors with large connections collaborate with different low degree interconnected communities to form a stable group.

## 4. Discussion

The organizing patterns of actors in national (Bollywood) and regional (Tamil and Malayalam) movie industry show high similarity in degree distribution, strength distribution, average weight as a function of end pint degree and weighted clustering coefficient. However we observe an important deviation in affinity measure which is indicated as average weighted nearest neighbour's degree as a function of degree. The regional movie exhibits dissassortative behaviour indicating its ability to absorb new actors. The Bollywood actor network is topologically similar to scientist collaboration network[1] (highly assortative and lack of clustering of hubs). The Tamil and Malayalam actor network is similar to knowledge network[3] (highly disassortative and lack of clustering of hubs).

Systems with co-occurring node interactions can be analyzed using weighted network measures. We have applied the same for analyzing movie actor network. In future we wish to increase our sample size and include more regional movie industries in our analysis.

## References

1. Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America, 101(11), 3747-3752.
2. Duvvuru, A., Kamarthi, S., & Sultornsanee, S. (2012, May). Undercovering research trends: Network analysis of keywords in scholarly articles. In*Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on* (pp. 265-270). IEEE.
3. Xiongfei, J. (2007). The Architecture of a Novel Weighted Network: Knowledge Network. *arXiv preprint arXiv:0711.4433*.