Structure and Evolution of Missed Collaborations in Large Networks

Minh X. Hoang
University of California, Santa Babara
mhoang@cs.ucsb.edu

Ram Ramanathan
Raytheon BBN Technologies
ramanath@bbn.com

Ambuj K. Singh
University of California, Santa Babara
ambuj@cs.ucsb.edu

***Abstract*—We study the nature of missed collaboration opportunities in evolving collaboration networks. We define a k-way missed collaboration as one in which every $(k-1)$-subset of the k persons has collaborated but the set of k has not. Representing a collaboration network as a *simplicial complex*, we model a missed collaboration as a *Minimal Non Face (MNF)*. Focusing on 2-dimensional and 3-dimensional MNFs, equivalent to 3-way and 4-way missed collaborations respectively, we analyze the DBLP publication network and the IMDB movie network.**

Our key findings are as follows. A large number of missed collaborations arise, but only a few persist for long. Specifically, the persistence time appears to be exponentially distributed for both 2-MNFs and 3-MNFs. Nodes with higher degree centrality are more likely to be part of 2-MNFs but little correlation was found with 3-MNFs. Considering the network of missed collaborations, the number of components as of year 2013 appears to be power law distributed across MNF types and data sets, but its evolution shows a divergence between DBLP and IMDB. Our results can help in developing random generative models of collaboration networks, cue researchers in on potential fruitful collaborations, and predict new collaborations.

I. Introduction

A social collaboration network is a set of *actors* (e.g. researchers) who interact with each other by means of certain *collaborative acts* (e.g. co-authored publications) [1]. The value of strong collaborations in making an impact cannot be questioned [2], [3]. Many collaboration networks are formed largely autonomously, without any centralized control on collaboration. In such networks, do all fruitful collaborations come to bear? Or are there collaborations that appear natural and potentially fruitful, but do not come to pass even after a large number of years?

We investigate the nature of such missed collaborations in large collaboration networks. A missed collaboration between a set of k actors is one in which a k-way collaboration would be meaningful, but did not happen. We use a purely structural way of determining "meaningful", based on the participants' existing collaborations. Specifically, we say that there exists a k-way missed collaboration if every $(k-1)$-cardinality subset of the k actors have collaborated, but the k actors as a set have not. For example, if (A,B), (B,C) and (C,A) have co-authored three different publications, but there is no publication where A, B, and C are joint co-authors, we say that there is a 3-way missed collaborations. In this example, the fact that A, B and C have pair-wise collaborated implies that a 3-way collaboration would likely be meaningful and useful, yet they have missed collaborating on a paper together. The concept can similarly be extended to 4-way, 5-way and general k-way missed collaborations.

Collaboration networks are traditionally modeled by graphs (e.g. [4], [17]). However, such a graph representation does not capture collaboration as a *group*. For example, consider a complete co-authorship graph on 4 vertices that represent four authors. Does this graph represent a single paper with four authors, four 3-author papers, or six 2-author papers? In particular, four 3-author papers is a missed 4-way collaboration, whereas a single 4-author paper is not, but this difference cannot be discerned using graphs. While some researchers have suggested the use of bi-partite graphs with edges between actor vertices and collaboration vertices, analysis is often done on a "one-mode projection" of these graphs [1]. What we really need is an abstraction where higher-order aggregations can be represented distinctly from the union of pair-wise collaborations.

In this paper, we use the *abstract simplicial complex* to represent and analyze collaboration networks. An abstract simplicial complex consists of a set V and a set of subsets of V closed under the subset operation. A simplicial complex is a generalization of a graph and therefore admits any analysis or metric based on graphs. Additionally it provides analytical possibilities not possible with a graph-based representation. In section II-A we provide a brief background on simplicial complexes as necessary for understanding this paper. Prior works [5], [6], [7] have established the usefulness of simplicial complexes for analyzing collaboration networks[1]. We capture a missed collaboration as a well known feature of simplicial complexes called a *Minimal Non Face* (MNF). In particular, a k-way missed collaboration translates into a $(k-1)$-MNF in the associated collaboration complex. We focus on 2-MNFs (3-way missed collaborations) and 3-MNFs (4-way missed

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053 (the ARL Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

[1]The *hypergraph* [8] is another possible abstraction, but as argued in [6], a simplicial complex is a better fit as it is closed under subsets, capturing the subset closure property of the collaboration relationship.

collaborations) as they are the only non-trivial MNFs occurring frequently enough for statistical analysis.

Our analysis has uncovered some key properties that cut across the two data sets. The persistence (number of years a missed collaboration lasts) is exponentially distributed for both MNF types. In other words, a vast majority of missed collaborations happen naturally in a few years after they form. The number of components of the MNF-induced network is power-law distributed for both MNF types. The number of MNFs a vertex is part of has a high correlation with the vertex and facet degrees for 2-MNFs, but not for 3-MNFs.

Some other features are remarkable in their differences across DBLP (publication database) and IMDB (movie database). Whereas MNFs grow exponentially with time in DBLP, we can discern no clear pattern for IMDB, and in fact, the MNF increase in IMDB is surprisingly not even monotonic. Further, the growth of the number of components in the MNF-induced network slows in the latter years for DBLP (indicating an increase in connectivity growth), whereas it is just the opposite for IMDB. We discuss possible explanations for these phenomena in section III-B.

Our statistical observations can be useful in constructing new generative models of evolving collaboration networks, and our technique for identifying specific missed collaborations can be useful in recommending potentially fruitful new collaborations. The observation that most MNFs have low persistence can help in models that predict future collaborations.

While this paper focuses on the application of MNFs to social collaborations, our techniques can be applied to communications networks as well. For instance, in a multi-channel multi-radio (MC-MR) ad hoc network modeled as a simplicial complex [9], an MNF represents opportunities for frequency re-assignment to increase broadcast efficiency. Cooperation between communications nodes is a key factor in spectrum sensing [10], cooperative transport [11], etc., and MNFs identify missed opportunities in such settings. Details can be found in [6].

II. Preliminaries

A. Simplicial Complex

***Definition** 1:* An *abstract simplicial complex*, or *simplicial complex* for brevity, is denoted by $\Delta = (V, S)$, where V is a set of vertices, and $S = \{S_i | S_i \subseteq V, S_j \in S\ \forall S_j \subseteq S_i\}$ is a non-empty set of subsets of V closed under the subset operation. $S_i \in S$ is called a *simplex* or a *face*. The dimension of a simplex S_i is $dim_{S_i} = |S_i| - 1$, and of the whole simplicial complex is $dim_\Delta = max\{dim_{S_i} | S_i \in S\}$. A *facet* is a maximal face F, such that $\{S_i \in S | F \subseteq S_i\} = \{F\}$. The *facet degree* of a vertex is the number of facets that the vertex is a part of.

Figure. 1 shows a simple example simplicial complex. The facets are $(0, 1, 2), (2, 3, 4)$, and $(1, 4, 5, 6)$, and the faces (simplices) are the subsets of the facets, including the facets themselves. The dimension of this simplicial complex is 3. The vertex degree of vertex 1 is 5, whereas its facet degree is 2. Obviously, a graph is a special case of a simplicial complex, i.e., a simplicial complex of dimension 1.

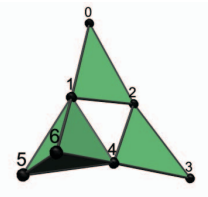


Fig. 1. Example simplicial complex

***Definition** 2:* A *minimal non-face* (MNF) in a simplicial complex $\Delta = (V, S)$ is a set of vertices $M \subseteq V$ such that $M \notin S$ and $S_i \in S\ \forall S_i \subset M$. If $|M| = k + 1$, it is called a *k-minimal non-face* (k-MNF), or an MNF of dimension k.

An 1-MNF is a missing edge and represents an uninteresting trivial missed collaboration. In Figure 1(a), $(1, 2, 4)$ is a 2-MNF. Consider the following complex: $\{(1, 2, 3), (2, 3, 4), (1, 2, 4), (1, 3, 4)\}$. In this complex, the set $(1, 2, 3, 4)$ is a 3-MNF, or a 4-way missed collaboration.

We have only given the minimum background required for understanding the rest of the paper. Readers interested in learning more about simplicial complexes and algebraic topology in general are referred to [12].

B. Datasets

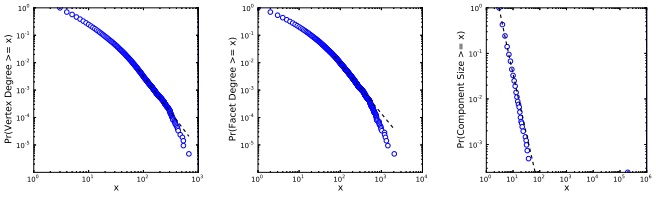
1) The DBLP Computer Science Bibliography: The DBLP Computer Science Bibliography is an online reference for bibliographic information on major computer science publications [13]. We extract all of the papers in this database from 1936 until September 2013 to create a dataset of 3,625,017 papers and 1,302,447 authors.

2) IMDB - The Internet Movie Databases: The Internet Movie Database (IMDB) is an online database of information related to films, television programs and other productions [14]. The database includes information regarding actors, actresses, directors, year of release, and other film-related information from year 1894 to 2013. We extract all films and the cast whose credits are less than or equal to 5 (the most important actors/actresses) to create a dataset with 488,238 cast members and 1,057,991 film titles.

When using graphs, each person can be represented as a vertex, and the collaboration between two people can be represented as an edge. The average vertex degree over time of the two dynamic graphs are shown in Figure. 6.

C. Representation as a simplicial complex

We represent each member in a data set as a vertex, and each collaborative act (a movie or a paper) as a simplex of vertices comprising it. Simplices may share vertices. Thus, in the DBLP complex, each vertex represents a researcher and each simplex represents a collaboration relationship among the researchers on one or more papers. Note that the number of facets may be less than the number of papers – for example, if there is a paper P_1 by (A, B, C), and P_2 by (A, B), we



(a) Vertex Degree (b) Facet Degree (c) Component count

Fig. 2. Complementary Cumulative Distribution Function (CCDF) for vertex degree (VD), facet degree (FD) and component size of the 2-MNF-induced networks (DBLP 2013), log-log scale

only have facet (A, B, C). The closure property of simplicial complexes means that the (A,B) collaboration is automatically part of the complex given (A,B,C), and does not have to be separately tracked (unlike in a hypergraph). Similarly, in the IMDB simplicial complex, each vertex represents a cast member and each simplex represents a collaboration relationship amongst the cast members on one or more productions.

The evolution of collaboration is captured as a cumulative simplicial complex. Let SC(y) represent the simplicial complex using data from only year y, and SC(1) the first available year complex. Then, the evolving simplicial complex EvoSC(i) = $\cup_{y=1}^{i}$ SC(y).

The *persistence* of an MNF M that first appears in EvoSC(j) is the number of consecutive years Y such that M is in EvoSC($j+k$) for every $0 \leq k \leq Y$, and M is not present in EvoSC($j+Y+1$). For example, if an MNF appears in the evolving 1943 complex and is present in 1944, 1945 and 1946, but not in 1947, then the persistence of the MNF is 3.

Some of our studies pertain to the network of MNFs. An *MNF network* or *MNF complex* of a simplicial complex has a vertex set equal to the union of vertices comprising the MNF and a facet corresponding to each MNF. In other words, it is the sub-complex induced by the MNFs of the complex.

III. Missed Collaborations: Structure and Evolution

We present a number of findings relating to the structure of MNFs in the DBLP and IMDB networks, and their evolution and persistence over time. Below, we use 2-MNF synonymously with a 3-way missed collaboration, and 3-MNF synonymously with a 4-way missed collaboration. The number of k-MNFs for k $>$ 3 are too few to draw conclusions from and therefore we do not consider them in our study. We thus have four combinations of {DBLP, IMDB} x {2-MNF, 3-MNF} studies in each section below. Note that, as stated in section II-C, the simplicial complex referred to for year i for all of the below is the *evolving* complex which contains all collaborations up to that year i.

A. Structure as of 2013

We consider the MNF complex in the DBLP and IMDB, and study the question: *What is the distribution of the vertex and facet degrees, and number of components? In particular, given the preponderance of power law in network science, are these also power law?*

TABLE I
Power-law fit for distributions of degree and component size $P[X = x] \propto x^{-\alpha}$ for $x \geq x_{min}$

	Feature	2-MNF induced network			3-MNF induced network		
		x_{min}	α	p-value	x_{min}	α	p-value
DBLP	Facet deg.	72	2.83	0.17	11	3.30	1
	Vertex deg.	79	3.63	0.97	9	5.19	0.98
	#Comps	3	3.55	0.82	8	3.01	0.93
IMDB	Facet deg.	440	2.69	0.04	118	2.51	0.68
	Vertex deg.	104	4.30	0.17	19	2.91	0.02
	#Comps	4	2.62	1.00	14	2.01	1.00

TABLE II
Pearson correlation of vertex degree and the number of MNFs that vertex belongs to (Year 2013).

		2-MNF	3-MNF
IMDB	FD	0.76	0.51
	VD	0.79	0.19
DBLP	FD	0.74	0.11
	VD	0.78	0.15

Figure 2 shows the distribution of vertex degree, facet degree and number of components for DBLP 2-MNF on a log-log scale. Visually, this appears to be power law distributed. However, visual analysis can be deceptive, hence, we analyze the distribution using Clauset's methodology [15]. The results are shown in Table I. A p-value < 0.05 rejects the power-law hypothesis, and a higher $xmin$ value dilutes it. From the table, it appears that power law is indicated as a good fit for component count distributions for all four combinations, and for vertex and facet degrees of DBLP-3-MNF. The facet degree of IMDB-2-MNF and vertex degree of IMDB-3-MNF do not follow a power law, while the remaining combinations show a somewhat weak fit to power law.

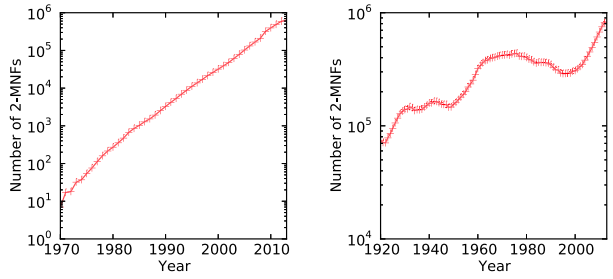
Thus, MNF network is structurally different in terms of degree distributions, with DBLP MNF networks having more of the well-known scale-free properties. However, surprisingly, from a global perspective of connectivity, they are similar, displaying strong power law properties.

Are higher degree vertices more likely to be part of more MNFs? Table II summarizes the Pearson correlation between the vertex/facet degree of a node and the number of MNFs it belongs to, for each of the four combinations. We observe that vertex and facet degrees for both IMDB and DBLP are correlated fairly strongly with the number of 2-MNFs, but only very weakly with the number of 3-MNFs. Thus, it appears that actors with high degree centrality are more at risk for missing 3-way collaborations. This is intuitive because the density around a node generates more collaborations overall and hence more missed, but the fact that this is not true for 4-way collaborations is somewhat surprising.

B. Evolutionary Characteristics

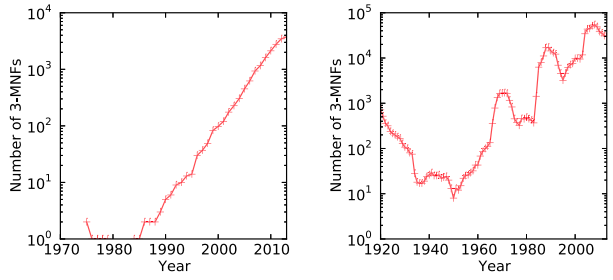
In this section, we study the following questions; *(a) How does the number of MNFs evolve over time?* and *(b) How does the connectivity of the induced MNF network vary over time?*

Figures 3 and 4 plot the number of MNFs as a function of years, in a semi-log plot. DBLP clearly shows exponential growth for both 2- and 3-MNFs. The 3-MNFs, not surprisingly, start appearing at a much later date due to the required



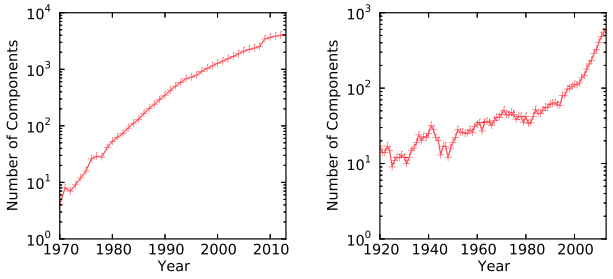
(a) DBLP (b) IMDB

Fig. 3. Number of 2-MNFs over the years, semi-log scale



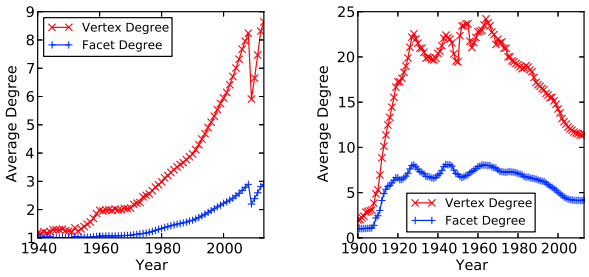
(a) DBLP (b) IMDB

Fig. 4. Number of 3-MNFs over the years, semi-log scale



(a) DBLP (b) IMDB

Fig. 5. Number of connected components of the 2-MNF induced network over the years, semi-log scale



(a) DBLP (b) IMDB

Fig. 6. Evolution of average degree

density and are fewer in number, but increase exponentially nonetheless.

On the other hand IMDB doesn't show a clear pattern, and in fact the number of 2- and 3-MNFs dips multiple times. This might have to do with the connectivity behavior of the original network (vice the MNF complex) as the MNFs can only form when there is good connectivity in a region.

Figure 5 plots the number of connected components as a function of the years for 2-MNF network. We observe an interesting divergence between the behavior in DBLP and IMDB – whereas the growth in the number of components tapers off for DBLP during the latter years, it actually increases for IMDB toward the latter years. The behavior for 3-MNFs is a more muted version of the same behavior, and not shown here for space constraints. We believe this might also be a direct consequence of the increasing and decreasing connectivity of original network for DBLP and IMDB respectively.

To test the hypothesis that the divergent behavior between DBLP and IMDB in the evolution of the number of MNFs and connected components could be attributable to the connectivity of the whole underlying complex, we examine the average vertex degree of the underlying networks, which generally correlates with connectivity (Figure 6). Indeed, we observe that the vertex degree of DBLP ramped up in the last few decades whereas that of IMDB sharply decreased. The "densification" phenomena over time has also been reported in [16] for publication and patent networks, supporting our hypothesis.

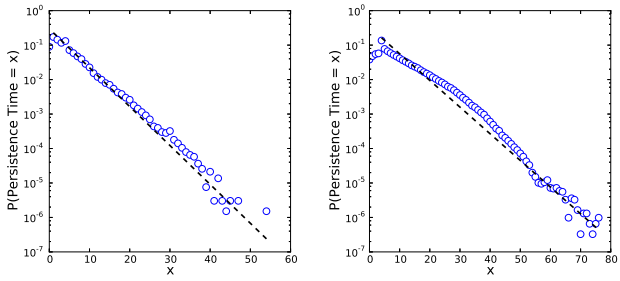
Why would the IMDB network get less dense with time? One reason for this could be that, unlike DBLP, IMDB consists of not only movies, but also documentaries and other productions, and from a number of different countries. Since each genre of production and each country tends to have its own largely disjoint community, and the diversity of genre and nationality has increased in the last few decades, the network is likely to have more components in the latter decades. Using only movies, and only from Hollywood could show smooth trends paralleling DBLP, but is left for future work.

We note that we have also generated the vertex- and edge-size normalized versions of the plots (not included here due to space constraints), but they do not show a remarkable difference in character.

C. Persistence Properties

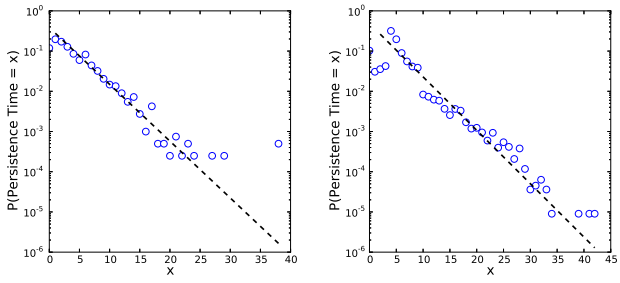
In this section, we study the question: *How long does a missed collaboration persist? What is the distribution of this persistence time?*

Based on the ubiquity of power law in network science, one might conjecture that the distribution might follow power law. Figure 7 and 8 show, respectively, the distributions of the persistence time of 2-MNFs and 3-MNFs plotted on a semi-log scale. The figures show that the persistence time of MNFs is not power law, but appears to be *exponentially distributed*. The fitting parameters for exponential distributions is shown in Table. III. This means that the majority of MNFs that arise get closed quite soon, and only a few tend to last very long. For example, 72% of 2-MNFs and 84% of 3-MNFs have a persistence time $\leq$ 5 years. IMDB 2-MNFs' average persistence time is much longer compared to DBLP, in particular 9.12 years vs. 4.37 years. This is likely because



(a) DBLP (b) IMDB

Fig. 7. Persistence time of 2-MNFs over the years



(a) DBLP (b) IMDB

Fig. 8. Persistence time of 3-MNFs over the years

IMDB history starts earlier. Additionally, the fact that most MNFs tend to close soon means that we can use MNFs to predict future collaborations.

To further investigate why MNF persistence time decays exponentially, assume that the probability for an MNF to remain open in a single time step is a constant p_{open}. Then the probability that it persists during x time steps and then closes would be $p_{open}^{x}(1-p_{open}) = (1-p_{open})e^{x \log p_{open}}$, which is indeed the form of an exponential distribution as fitted in Table. III and shown as dashed black lines in Figure 7 and 8, where $\lambda = -\log p_{open}$. In addition, p_{open} is smaller for 3-MNF networks than that for 2-MNF networks (refer Table. III), which suggests that 3-MNFs are more easily closed compared to 2-MNFs. One possible explanation could be that the history of existing collaborations is tighter in 3-MNFs (all pairs plus all triplets) in comparison to 2-MNFs (all pairs).

We note that a k-MNF by definition pre-selects nodes that have collaborated pairwise, so there isn't a fundamental difference of opinions or animosity between any two actors that might be a barrier. Thus, all things being equal, the nodes in the k-MNF are more likely to naturally collaborate. We would thus expect p_{open} for MNFs to be smaller than for other subsets of the same cardinality. Indeed, for DBLP, p_{open} for a random subset of three nodes is very nearly 1.0. Even if two nodes are chosen such that they are at most two (one) hops from a given node, the p_{open} is 0.98 (0.79), attesting to the relative value of MNFs in predicting future collaborations.

Unlike the two previous studies in sections III-A and III-B, there is no significant difference in the behavior of MNF persistence between DBLP and IMDB. Thus it appears that

TABLE III
EXPONENTIAL FIT FOR DISTRIBUTIONS OF PERSISTENCE TIME
$P(X = x) \propto e^{-\lambda x}\ \forall x \geq x_{min}; p_{open} = e^{-\lambda}$

	2-MNF			3-MNF		
	x_{min}	λ	p_{open}	x_{min}	λ	p_{open}
DBLP	1	0.26	0.77	1	0.31	0.73
IMDB	4	0.18	0.84	2	0.31	0.73

TABLE IV
SUMMARY OF OBSERVATIONS

	2-MNF	3-MNF
DBLP	**#Comps is strongly power law** VD/FD weak power law High correlation w/ VD/FD *#MNFs grow exponentially* *#Comps growth slows in tail* **Persistence exponentially distr**	**#Comps is strongly power law** VD/FD power law No/weak correlation w/ VD/FD *#MNFs grow exponentially* *#Comps slows in tail* **Persistence exponentially distr.**
IMDB	**#Comps is strongly power law** VD/FD not power law High correlation w/ VD/FD *#MNFs growth non-monotonic* *#Comps growth faster in tail* **Persistence exponentially distr.**	**#Comps is strongly power law** VD/FD weak power law No/weak correlation w/ VD/FD *#MNFs growth non-monotonic* *#Comps growth faster in tail* **Persistence exponentially distr.**

insofar as individual MNF evolutionary features are concerned (Figures 7 and 8), there is more uniformity, whereas if network-wide features are concerned (Figures 3, 4, and 5), there is a marked difference. This is probably related to the difference in the whole (vice MNF-induced) network.

D. Discussion

Table IV summarizes the observations from the previous subsections in the order they were presented. The observations that hold across all four combinations of data sets and MNF types are in bold. Those that differ markedly across DBLP and IMDB are shown in italics. Within each of DBLP and IMDB, other than the correlation of vertex and facet degrees to number of MNFs, observations largely hold across 2-MNFs and 3-MNFs.

A key finding that holds across both data sets and MNF types is that most MNFs close within a few years, and far sooner than other equal-sized actor sets. The implications are two fold: (1) the fact that most of them close naturally means that MNFs do identify valid missed collaborations for the most part; (2) an MNF created in a particular year predicts collaborations that are likely to occur in the next few years, namely, the collaborations represented by the MNFs. The differences between DBLP and IMDB – e.g. in the evolution of number of MNFs and the components in the induced MNF network – may be explained to some extent by the nature of the communities: Unlike DBLP which is largely homogeneous, IMDB, having not only movies but also documentaries and TV programs, and from diverse countries, is more disconnected.

Some caveats are in order in interpreting our results. First, there is a tacit assumption that the presence of all possible (k-1)-way collaborations implies that a k-way collaboration makes sense. This may not be true in some specific circumstances. In such cases, the MNFs will be "false positives", i.e., collaborations that upon further examination are not meaningful. Addressing this is beyond the scope of this paper. Second, the identified missed collaborations may have

happened outside of DBLP or IMDB. Third, the names of the same person may differ, or different people may have the same name in the database. Finally, our trimming of the IMDB database by considering only the top credited actors/actresses may not be a representative sample of the whole.

IV. Related Work

The last decade has seen a spurt in the study of collaboration networks [4], [17], [1], [18], [19]. In [17], [4], structural properties of publicly available scientific publication networks are analyzed, with the latter also investigating evolutionary aspects. Movie actor and the DBLP networks are analyzed in [19]. Self-organization and classification into different kinds of small-world networks appear in [1], [19]. The quest to discern power laws in the distribution of social networks has been the subject of much work in the literature [15], [20]. Visual and other simplistic methods, however, may be misleading [15]. A rigorous method for verifying if a distribution follows power law is given in [15], which we apply in this paper.

Well established in mathematics, in particular algebraic topology [12], simplicial complexes have been used as a part of Q-analysis in the 1970s to analyze general structure [5], and have been applied into specific social network problems [21].

The application of simplicial concepts to collaboration networks appears to varying extents in [6], [7]. In [7], 2-MNFs were briefly studied, but for much smaller collaboration networks. There has been some work on recommendation systems for collaboration (e.g. [22], [23], [24]), but these works are focused on recommending one other individual, need profile information, do not consider the evolutionary information, and do not study the statistical properties.

To our knowledge, ours is the first work that investigates the structure and evolution of multi-way missed collaborations in large, autonomous networks using a simplicial model.

V. Concluding Remarks

We have studied the structure and evolution of missed collaborations in DBLP and IMDB by modeling it as a Minimal Non Face (MNF) in the corresponding simplicial complex. We have discovered that some properties appear to have more general validity – e.g. distribution of MNF persistence time; and some properties are more specific to the data set as a result of unique characteristics of that data set – e.g. connectivity evolution. Our statistical results can be used to create or validate random generative models tailored to the nature of the particular network type, and our techniques for identifying missed collaborations, along with the property that they are less likely to remain open compared to other subsets, can be used as a recommender system for multi-way collaborations, or predict new collaborations.

There are a number of interesting avenues for future work: a more in-depth analysis of the results, and better support for our explanations; further substantiation by analyzing other data sets (e.g. ArXiv, PubMed, SourceForge); an efficient generalized algorithm for computing k-MNFs and persistence across time; relation with other metrics (e.g. clustering coefficient); and relationships with other features of simplicial complex, for e.g Betti numbers.

References

[1] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras, "Self-organization of collaboration networks," *Phys. Rev. E*, vol. 70, Sep 2004.

[2] J. D. Adams, G. C. Black, J. R. Clemmons, and P. E. Stephan, "Scientific teams and institutional collaborations: Evidence from us universities, 1981–1999," *Research Policy*, vol. 34, no. 3, pp. 259–285, 2005.

[3] S. Wuchty, B. F. Jones, and B. Uzzi, "The increasing dominance of teams in production of knowledge," *Science*, vol. 316, no. 5827, pp. 1036–1039, 2007.

[4] A. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A: Statistical Mech. and its Applications*, vol. 311, no. 3, pp. 590–614, 2002.

[5] R. Atkin, *Mathematical structure in human affairs*. Heinemann Educational London, 1974.

[6] R. Ramanathan, A. Bar-Noy, P. Basu, M. Johnson, W. Ren, A. Swami, and Q. Zhao, "Beyond graphs: Capturing groups in networks," in *Proc. IEEE NetSciComm*, pp. 870–875, IEEE, 2011.

[7] T. Moore, R. Drost, P. Basu, R. Ramanathan, and A. Swami, "Analyzing collaboration networks using simplicial complexes: A case study," in *Proc. IEEE NetSciComm*, pp. 238–243, IEEE, 2012.

[8] C. Berge, *Hypergraphs*. North-Holland, 1989.

[9] W. Ren, Q. Zhao, R. Ramanathan, J. Gao, A. Swami, A. Bar-Noy, M. Johnson, and P. Basu, "Broadcasting in multi-radio multi-channel wireless networks using simplicial complexes," in *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference*, pp. 660–665, IEEE, 2011.

[10] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116 –130, 2009.

[11] R. Ramanathan, "Challenges: A radically new architecture for next generation mobile ad hoc networks," in *Proc. of ACM MobiCom*, (Cologne, Germany), Aug. 2005.

[12] E. Spanier, *Algebraic topology*, vol. 55. Springer, 1994.

[13] "DBLP computer science bibliography, http://www.informatik.uni-trier.de/∼ley/db/."

[14] "Internet movie database (IMDB), http://www.imdb.com/."

[15] A. Clauset, C. Shalizi, and M. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[16] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proc. of the 11th ACM SIGKDD*, pp. 177–187, ACM, 2005.

[17] M. E. J. Newman, "The structure of scientific collaboration networks," *Proc. National Academy of Sciences*, vol. 98, no. 2, pp. 404–409, 2001.

[18] P. Zhang, K. Chen, Y. He, T. Zhou, B. Su, Y. Jin, Y. Chang, H. an d Zhou, L. Sun, B. Wang, *et al.*, "Model and empirical study on some collaboration networks," *Physica A: Statistical Mechanics and its applications*, vol. 360, no. 2, pp. 599–616, 2006.

[19] L. Amaral, A. Scala, M. Barthélémy, and H. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11149–11152, 2000.

[20] M. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[21] P. Gould and A. Gatrell, "A structural analysis of a game: the Liverpool v Manchester United Cup Final of 1977," *Social Networks*, vol. 2, no. 3, pp. 253–273, 1980.

[22] G. R. Lopes, M. M. Moro, L. K. Wives, and J. P. M. De Oliveira, "Collaboration recommendation on academic social networks," in *Advances in Conceptual Modeling–Applications and Challenges*, pp. 190–199, Springer, 2010.

[23] H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles, "Collabseer: A search engine for collaboration discovery," in *Proc. of the 11th International ACM/IEEE Joint Conference on Digital Libraries*, pp. 231–240, 2011.

[24] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," in *Proc. of the 18th ACM SIGKDD*, KDD '12, pp. 1285–1293, 2012.