Triadic Closure in the IMDB

Evidence for its Existence and Impact

Rui Jin
Department of Science
Rensselaer Polytechnic Institute
Troy, NY, 10930, United States
jinr@rpi.edu

***Abstract*—Social networks have certain properties that tend to influence the formation of relationships. Triadic closure is one of these properties, and has a tendency to exist in real social networks. Specifically, if three people have two relationships among one another, there will be an increased chance of the third relationship forming. However, its reliability in actual networks may be questionable. By examining an example social network, the Internet Movie Database, it is possible to look for the existence of triadic closure, and its potential impact on a network.**

Keywords-social networks, triadic closure, IMDB, Internet Movie Database

I. Introduction

Relationships can form in many different ways. Any two given people might interact due to common interests or common purposes. However, even with certain shared traits, there is no guarantee that any given pair of people will form a relationship.

However, triadic closure can be used as a means of predicting certain relationship. From a straightforward perspective, it would be possible to see that if there are three people, and one person is friends with both of the others, then the third potential friendship will be more likely to form than in a completely random setting. Evidence of this has been found in the past, but there is a limited amount of confirmation. As a result, it would be beneficial is to gather evidence for triadic closure within the Internet Movie Database, and by extension, social networks in general.

While evidence of triadic closure is helpful, it is also important to find details of its specific implementation. By analyzing the precise impact of triadic closure, it will be possible to better gauge the accuracy of using triadic closure as a tool to predict relationship formation. The best evidence available would exist if the clustering coefficient of the graph was larger than that of a random graph. This would prove that there is a greater probability of triangles coming into existence than what would come into existence if triadic closure did not exist.

However, if the evidence for triadic closure is not clear, new information might still be drawn regarding the reasons for its existence or nonexistence. it is important to note that according to [1], there are three primary reasons for triadic closure. Trust, opportunity, and incentive may all be driving forces behind triadic closure. However, this also grants some insight into why triadic closure may not have a large impact in the IMDB, if any of these conditions would not be relevant.

In order to offer evidence for triadic closure, it is important to distinguish between weak and strong links. While triadic closure may exist without links of specific strength, additional data can be found if a distinction is made between weak and strong links. As a result, some consideration is given into the relative strength of any link, and is therefore evaluated when triadic closure is considered.

Finally, it is important to consider the specific details of the Internet Movie Database, and how its relationship formation mechanisms may influence the effect of triadic closure. While there are many models of graph creation, it can be difficult to properly simulate the creation and distribution of a graph based on movies. There are many reasons for this, but it can be assumed that movies have more barriers to creation in comparison to other means of relationship measurement, such as papers or “friend” additions on internet sites.

In order to fully analyze the information within this data set, this research has been divided into four major parts. The first part of the data involves a cursory glance at the data, in order to help grant a basic understanding of the IMDB and to help make judgments about the proceeding experiments. The second part uses that information to help make basic comparisons against randomly generated data. Using these refined graphs, it will be possible to compare this actual graph to randomly generated graphs with similar numbers of nodes and edges. Once all of the data has been properly analyzed and compared, then it will be possible to make conclusions based on traits such as the clustering coefficient of the graph. The third part of this paper focuses upon examining the IMDB in order to show the time necessary for triadic closure's effect to show. The fourth part of this analysis will be a cursory glance at the accuracy of triadic closure, and how well it is able to predict new social links. By analyzing this data, it will be possible to draw conclusions regarding the likeliness of triadic closure and the time necessary to close.

Through all of these experiments, it is possible to gather an amount of evidence for triadic closure, as well as prove that triadic closure generally needs a short time to show its effect.

However, its use as a predictor of future links is somewhat limited, and it is therefore important to properly consider its value in any graph before it is used as a tool.

II. Preexisting Research

Other papers and other research into the realm of graphs and triadic closure exist. However, the amount of research focused on triadic closure is limited, and none could be found specifically linking it and the IMDB.

Reference [2] speaks of a specific variation of triadic closure to consider. It primarily focuses on the Twitter network, and how closure might be represented in a directed network. While this research is focused upon a directed network, its insight into undirected networks and the strength of ties help to predict various possible occurrences. It also goes over the potential effect of shared interests.

Reference [3] focuses more on the formation of online social networks, and mentions various different influences on graph creation over time. While it does not specifically focus on triadic closure, its insight into models involving more than preferential attachment help to consider other potential reasons for node creation.

Reference [4] provides an overview of small world problems, where the general degree of every node is small. It also explains why movies can be modeled as small-world. As a result, it provides various useful facts to use when considering how to properly model and analyze the data set.

Reference [5] provides insight into the preferential attachment model, and offers explanations for how preferential attachment models of graphs will result in the creation of small-world graphs. It helps provide insight to the creation of a random graph which would could help provide more information and another comparison to the actual graph of the IMDB data set.

Reference [6] helps grant a general visualization of the IMDB, including its general structure, creation over time, and a general cluster of the co-actor network. However, while this data grants general insights into the IMDB, it does not speak much about triadic closure within the graph.

Reference [7] helps to explain various means of finding connectivity within graphs. As a result, it is a useful tool in order to properly detect the potential number of triangles that may exist for the purposes of clustering coefficient.

There is a large amount of information available in regards to social network graphs as whole. However, the IMDB is an interesting example, and specific research regarding triadic closure can be difficult to find. It is hoped that this peculiar combination of data set and research focus will help to bring some new insight into the mechanisms behind triadic closure, as well as behind this special type of social network.

III. IMDB Initial Data

A. Initial Data Conversion and Analysis

The primary data set interpreted here is a small section of the Internet Movie Database [8]. Specific data was gathered in order to help fulfill the requirements necessary to work with a reasonable sample size, while retaining a large amount of information. Data is originally taken directly from the IMDB website, and the movie name, release date, and up to three actors involved in each movie are gathered within the data set for storage.

In order to interpret this data, this is converted from a text format to a graph, using the JGraph library [9]. Nodes and edges are added to the graph on a movie-by-movie basis.

Steps are taken to insert each actor and movie into the graph. First, if an actor does not already exist within the graph, they are added as a node. Then, each potential pair of actors within the movie are linked to one another. Finally, new edges have their weights set to 1, while existing nodes have their weights incremented by one. Using this method, it is possible to keep a running tally of the total number of movies that any two actors share.

By doing all of this, it is possible to get a basic overview of a graph for the IMDB data set. However, it is important to note that this will only result in a basic graph of the data. After these steps, there should be a resulting graph holding all of the necessary information to do a basic analysis of the Internet Movie Database. The focus on this initial analysis will be to attempt to identify important points which may act as good definitions for a social link. In particular, we are looking at how many shared movies to require for a weak link, and how many shared movies should be necessary for a strong link. In order to get this information, we will look at the total number of actors with relationships, and the number of relationships per actor that result due to a variety of requirements.

B. Initial Results

TABLE I. Actors with Relationships Based on Required Movies for Relationship

Number of Movies for a Relationship	*Number of Actors with Relationships*	*Average Number of Relationships per Actor*
1	404006	8.87
2	56859	5.76
3	19140	4.61
4	10246	4.15
5	6519	3.99
6	4778	3.78
7	3705	3.63
8	2961	3.52
9	2455	3.44

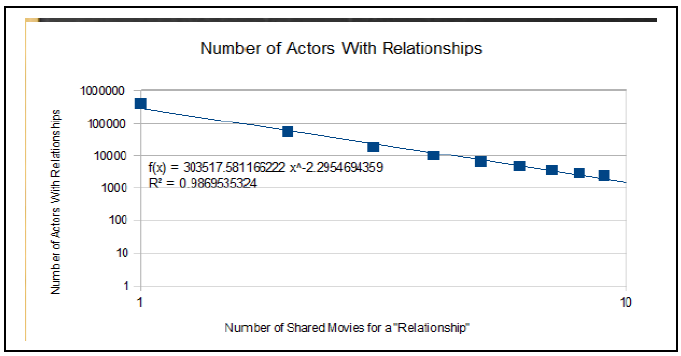

Figure 1. Actors v. Relationship Requirement on Log-Log Plot

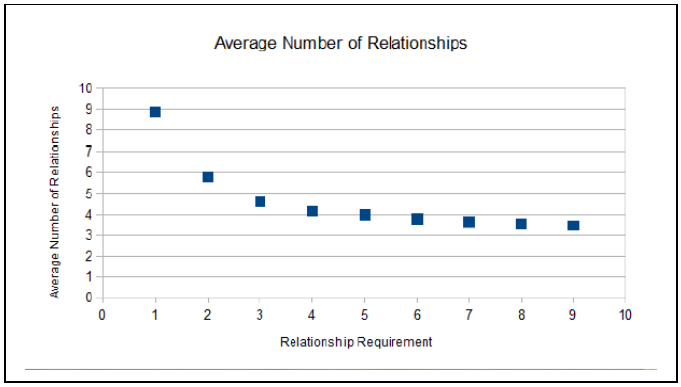

Figure 2. Relationship Requirement v. Average Number of Relationships Per Actor

While other data regarding these relationships can be drawn, the initial data is focused on analyzing the connection between the relationship requirement and the average number of relationships per actor in order to find good cut-off points for the definitions of weak and strong graphs. In particular, it is easy to note that, as seen in Fig. 2, there is an extremely sharp drop-off in the average number of relationships a person has when the relationship requirement is increased from one movie to two movies. In addition, it is possible to see that the decrease in average number of relationships quickly slows down.

C. *Initial Discussion*

From this data, it is possible to see that there is an extremely sharp drop between a requirement of a single movie, and the requirement of two or more movies. The number of relationships per actor suddenly drops from an average of about nine relationships per actor to about six relationships per actor. In addition, while there are further drops, these drops drastically decrease in size, as the requirement increases.

As a result, we will make the assumption that a weak link can be defined as a link in which two actors have been in exactly two shared movies. This number is chosen due to the extreme drop in the average number of movies that occurs after restricting the requirement from one movie to two. This also helps to confirm an actual relationship. Single shared movies could potentially be qualified as coincidental meetings, or one-shot working relationships which are broken afterward. By requiring two movies, it is possible to more accurately represent data beyond coincidence.

In addition, the requirement for a strong link will be set to three or more movies. Due to the slowing drop in the average number of movies, three movies will be an acceptable cut-off point, after which it becomes much more difficult to find a sudden sharp drop which would be indicative of a change in link strength. This is to take advantage of the last sharp drop in number of relationships before the decrease shrinks and becomes difficult to draw straightforward conclusions from.

Beyond these specific decisions, it is possible to see the basic ideas of social network graphs in play. It is easy to see that as requirements for a relationship increase, fewer people have them. In addition, fewer people actually have relationships if the requirement is set at a sufficiently high level. This simply grants general insight into how social networks can be modeled, and how it can be difficult to properly import their data to analyze in a meaningful way.

IV. IMDB Refined Data V. Random Data

A. *Data Conversion*

In order to get more specific data, it is necessary to reduce the full data set from earlier to more refined data. While it is possible to perform basic analysis upon the large data set, this would leave out possible insights into the full data set. As a result, it is important remove any data that would be unnecessary and could potentially interfere with clear results.

The first step involves removing irrelevant. The most important data to remove is the set of nodes which have no edges. There may have been created, due to movies which may only have a single actor. As a result, it is possible to have actors which do not share any edges. They would be exceptions to the general analysis, and could potentially throw data off.

In addition, it is possible to refine the data by identifying weak and strong links. Since the definition of a weak link has been set as a link of two movies, and the definition of a strong link as been set as a link of at least three movies, it is now possible to identify weak and strong links in the graph. Future analysis will therefore distinguish between the two types whenever weights are considered. It is also important to remove all of the edges which only have a weight of one, since that would indicate a given link does not have enough weight even to qualify as a weak link. Due to the selected definitions for weak and strong edges, these edges will not be considered in the overall analysis.

It should be noted that while the IMDB holds a link's strength in the edge weight, it does not hold information regarding whether a relationship is positive or negative. This is due to a lack of information and a lack of means to distinguish between the two in this data set. There is no practical way to determine whether or not any pair of actors have a friendly relationship or a rivalry. As a result, these investigations on triadic closure will be focused towards the strength of a relationship, and not the nature. They will simply work under the assumption that all links are positive.

B. Simulating Random Charts

In order to properly compare this data and confirm triadic closure, we must first generate random graphs which can be used. The three main graphs that will be considered are an Erdos-Reyni random graph, a random graph with its degree distribution approximated to that of the IMDB data set, and a Barabási-Albert graph.

The Erdos-Reyni graph is simply simulated by simulating the total number of nodes from the graph, and considering every potential edge pairing as a probability, based on the total number of edges possible and the potential number of edges based on the number of nodes in the graph [10]. By using this particular creation mechanism, and feeding in the number of nodes and edges in the actual IMDB data set, it is possible to simulate the average degree distribution of our data set. This allows for one potential comparison.

Another potential comparison can be made through a random graph with an approximated degree distribution. This is done by simulating a graph with a given amount of nodes, and passing in a target degree distribution for every node. The graph will then create a random list with one mark for every desired link of the given node, and then will go through sequentially in an attempt to match nodes until their degree requirements are fulfilled. This will result in a random graph with degrees similar to that of a real graph. It is important to note that this graph generation mechanism has a tendency to generate fewer edges than given. However, this is a good approximation for this type of graph.

Finally, a preferential attachment model might be a potential candidate for a random graph to compare to. However, there are several factors that result in potential issues with this random graph model. In particular, it is extremely difficult to properly weight this preferential attachment model without factoring for several oddities within this specific type of social network. In particular, the barriers to forming a link via movies are much higher than than the barriers to forming a simple link through social networks.

In addition, it is extremely difficult to provide an effective probability of preferential attachment being used, as opposed to random generation. As a result, the Barabási-Albert graph model has some practical difficulties within its implementation. This is due to the nature of the graph, where new links can be created between multiple new nodes at the same time. Because of its extremely focus towards the earlier nodes in the graph, attempts at finding triangles are extremely expensive computationally.

While this graph was attempted with a probability of 0.5, it was found that attempts at trying to compute the clustering coefficient of such a graph was computationally expensive, due to the extreme number of triangles that existed within the focused nodes of a graph.

C. Triadic Closure Analysis

With this data available, it is important to begin looking at triadic closure in particular. After creating each of the potential graphs for all links, as well as considering the graph of only strong links, we will find the total number of edges for each of the graph, the total number of triangles that are completely formed, and the average local clustering coefficient for the full graph. It is important to note that while it is possible to find global clustering coefficient as well, but it has been foregone in this case due to technical difficulties. As a result, a focus will be placed on local data, rather than global data. In addition, due to the nature of a Barabási-Albert graph, even local clustering coefficient can take a large amount of time to compute.

There are a few important details to note for these decisions. The total number of edges was recorder in order to help make proper comparisons with the IMDB. The number of vertexes were kept constant simply because every random graph was created with the number of vertexes in mind. From there, the number of triangles was counted without regard to edge strength, and the local clustering coefficient was averaged for every individual node.

D. Results

In general, the most valuable results from these graphs are the clustering coefficients. In order to get a good grasp of all of the information, below are the number of triangles for graphs when considering both strong and weak links within triangles, as well as only considering strong links.

TABLE II. COMPARISON OF IMDB TO RANDOM GRAPHS – WEAK AND STRONG LINKS

Type of Graph	*Total Number of Edges*	*Total Number of Triangles*	*Average Local Clustering Coefficient*
IMDB	163482	1014056	0.67
Erdos-Reyni	163077	312161	0.35
Random With Degree Imitation	138639	135187	0.25
Barabási-Albert	170538	219144	0.09

TABLE III. COMPARISON OF IMDB TO RANDOM GRAPHS – STRONG LINKS ONLY

Type of Graph	*Total Number of Edges*	*Total Number of Triangles*	*Average Clustering Coefficient*
IMDB	44148	212814	0.67
Erdos-Reyni	44164	67891	0.37
Random With Degree Imitation	41191	34538	0.25
Barabási-Albert	38265	23716	0.06

The information from these tables show that the IMDB data set has a much greater number of triangles and a much higher clustering coefficient than would exist for a similar random triangle. As a result, it is safe to say that the number of triangles that result are obviously not random, and that triadic closure does exist. However, it is important to note that we can see the average clustering coefficient of each node is

significantly higher in the IMDB, in comparison to that of the others.

It would be natural that the IMDB has a comparatively high number of triangles, in comparison to the randomly generated graphs. This indicates that despite the same number of nodes and similar number of edges, there is more of a tendency for the graph to form clusters and triangles, in comparison to random graphs.

In addition, the average clustering coefficient for the Barabási-Albert graph was exceedingly low, and as a result, the data could not be considered useful with large graphs. Note that while the specific implementation of this graph may have had issues with randomness in the total number of edges, this should not greatly influence the overall clustering coefficient. As a result, the Barabási-Albert model is not recommended in future research within the IMDB Database. While it is possible to use it as a model, the computational cost for doing so is extremely high for any situation where the probability of preferring busy actors is high enough to have a notable difference from that of a random node.

While a full confirmation of global clustering coefficient is not available, this data has a good amount of evidence towards triadic closure. The partial evidence through the total triangle count and average clustering coefficients both offer evidence towards triadic closure.

V. IMDB Triangles Over Time

A. Data Set Conversion

In order to analyze the effect of triadic closure over time, it will be necessary to add the year of movie release into the compete graph. However, it could get computationally expensive to analyze all movie release dates, as well as dynamically figure out when a specific relationship is created. It might also be difficult in distinguishing between weak links and strong links during the exact moment of closure. Therefore, this paper will work with the years of initial meetings, as opposed to keeping track of the year of release for every individual movie.

As a result, this paper will work under a simplified assumption that it takes no time for a relationship to fully form into its target. While this type of assumption is not completely accurate, it reduces the computational power necessary to get results, and will allow for a fair amount of insight into the overall graph design.

It is important to note that from this stage on, triangles will only be considered if there is at least one strong link within the triangle. While this was previously unnecessary due to a lack of distinction between weak and strong links in random graphs, it is now possible to draw a distinction between triangles which are formed with only weak links, and triangles which include strong links.

The current definition of triadic closure only deals with triangles which will close due to unstable relationship involving at least one strong link. As a result, it is now possible to add this restriction to the list of considerations. Because of this, the number of triangles found below will differ from the previous data, due to the large number of triangles which only exist as a result of weak links.

B. Data Gathering

In order to focus on the time it takes for triangles to close, this data set will map the number of years it takes for any given triangle to close. In order to do this, it searches for any strong link in a triangle. (According to [11], If a triangle does not have a strong link, then it cannot properly close.) Once a strong link is found, all potential triangles including that link are considered, and counted if a pair of weak links exist. Finally, once all triangles with that link are accounted for, that link is removed from the graph to prevent repeats from being counted. This will result in a count of the total number of triangles necessary.

In order to gather the years of time necessary for each triangle to form, the years are simply ordered among the three potential edges in a triangle. The second year is always subtracted from the third, indicating the time that the triangle had before being unstable, and the time it had when the triangle fully formed.

C. Results

The results of this data can be seen in Table IV. While there are more data points beyond the 35 year mark, that data has been discarded due to being too scattered to draw reliable conclusions on. However, it is important to note that the existence of such data shows that while the probability of closing a triangle decreases with time, it does not fully vanish for a large number of years.

TABLE IV. Time for Triangles to Close

Number of Years to Close	***Number of Triangles Closed Within Timespan***
0	79265
1	29668
2	12196
3	6817
4	4484
5	3045
6	2184
7	1587
8	1152
9	889
10	711
11	527
12	426
13	328
14	266
15	190

Number of Years to Close	*Number of Triangles Closed Within Timespan*
0	79265
16	164
17	143
18	121
19	101
20	67
21	41
22	46
23	35
24	21
25	20
26	14
27	19
28	9
29	10
30	10
31	8
32	4
33	3
34	6
35	2

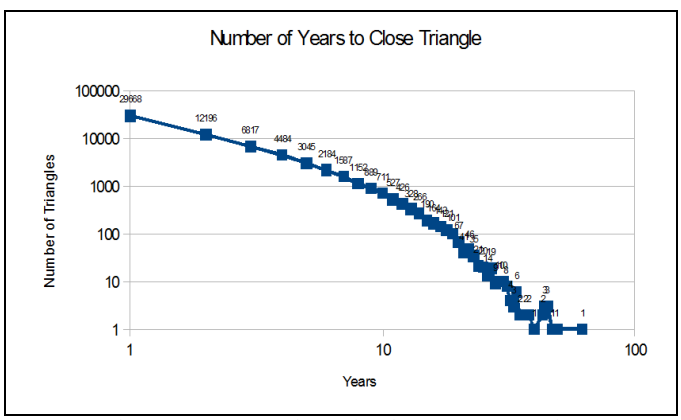

Figure 3. Years to Close Triangles on a Log-Log Plot

D. *Discussion*

From Table IV, it is possible to see that there is an extremely large slant towards immediate or near-immediate closure of triangles. Over half of all potential relationships that are formed close within the same year as the year an unstable triplet is created.

Because of the nature of this data, it is therefore safe to assume that actors tend to form relationships with the first people that they make a movie with. Due to the time it takes for any given movie to go from production to full release, it is essential to note that there will be time gaps between any two given releases.

As a result, the sheer number of zero-year triangle closures show that there is a large number of relationships that are started by any person's first movies, and that they will have a great tendency to continue working with the same people, as seen by the requirement for a strong relationship.

We can also see from Fig. 3 that at around 25 years, there is a slight sharp drop in the amount of time it takes for a triangle to completely close. While this slight drop may be due to coincidence, further research might be possible in order to find out more. There is a potential for a generational shift during that time period, which may explain the sudden difficulty in completing relationships beyond that point.

Despite the small gaps that may exist, it is important to note that the most clear conclusion from this data is that triangles tend to close quickly. While it is possible for triangles to close over time, the numbers of triangles which close over time tend to decrease rapidly as the years pass.

VI. Total Likeliness to Close

A. *Data Conversion and Gathering*

Very little conversion is necessary from the previous data set in order to interpret strong and weak bonds, and use them in judging likeliness to close. However, instead of pulling data out in regards to the time it takes to close, this focus will instead be on the chance of closure. As a result, only simple data is drawn from this information, primarily focused on the number of potential triangles in comparison to the number of unstable triangles which do not close.

In order to look for the likeliness of any unstable triangle to close, we will consider unstable triplets and complete triangles. It is important to note that in this situation, an unstable triangle is defined as a triplet of nodes with either a weak link and a strong link, and two weak links. Also of note is that if a triangle is complete, then the order of triangle completion is not considered. For example, if two weak links existed in a triplet of nodes, and a third strong link completed the triangle, this would register as a legitimate triangle.

The reason for this is primarily simplification. While it would be possible to check and confirm all of the dates, the exceptions to triangle formation would be small enough that the data as a whole should not be influenced. In addition, the data that might be pulled without this simplification would not be of much value, due to the previous simplifications made in regards to how the year of any given link is set.

For an unstable triangle to be counted, it must have at least one strong link within its two potential links. As a result, only count triplets with either one strong link and one weak link or two strong links.

B. Results

In general, it is possible to see that there are 1105340 unstable triangles which we would completely expect to close, when all triplets are considered. However, as tallied from the data earlier, there are a total of 144595 actual triangles. As a result, it is possible to see that only roughly 13.1% of all potential triangles are closed.

C. Discussion

Obviously, the extremely low percentage of closed triangles indicates that while triadic closure is an accurate method to predict a potential triangle. Triadic closure does provide a better measure of triangles than complete randomness, but it is obviously not an effective means of completely accurate prediction. This needs to be compared to a completely random graph, as well as other potential predictors of link formation. However, without any alternative means of attempting to properly predict the development of links over time.

As a result, while triadic closure clearly exists, it is difficult to successfully use and analyze it for link prediction. If only roughly 13% of all predicted triangles will actually close, then it is clear that the chance of successfully predicting a link using triadic closure are relatively low.

More research must be applied to look into triadic closure. However, it would also be beneficial to investigate the converse. That is, it would be beneficial to see what percentage of new links are due to triadic closure, and what percent of links are created for other reasons. Homophily and chance would play a role in such an experiment. With our specific graph, it is difficult to separate the reasons for new links. In addition, due to the simultaneous creation of multiple nodes, it can be difficult to detect whether or not a relationship is created due to triadic closure, or for other reasons. Due to the selected means of timing relationship creation, it is safe to say that people may already be interested in continued work with the same people. As a result, it can be difficult to detect exactly when the members of an unstable triangle are introduced to one another.

VII. Summary

From all of the data available, it is possible to draw multiple conclusions in regards to triadic closure.

1) Triadic closure most likely exists. This can be determined by the comparatively high clustering coefficient, when comparing the actual IMDB data set to a randomly generated data set.

2) Triadic closure generally does not need a long time to come into effect. The number of closed triangles tend to die down over time. Over half of the triangles were closed within the same year, and a total of over 80% of all of the triangles were closed within two years after that. As a result, it is safe to say that if triadic closure will occur, it will generally occur quickly.

3) An unstable triangle does not guarantee triadic closure. A low percent of unstable triangles were actually closed, despite the short amount of time it usually takes to close the triangle.

The inaccuracies that exist within these graphs might be due to a variety of reasons. For example, while triadic closure may normally be due to opportunity, opportunities to make new relationships through movies may be more difficult to come by, even with friends helping out.

In addition, further insight into the specific traits of the movies may provide further insights into the general form of this graph. Things such as sequels, the genres of the movie, and age of the actors may all play a role in how triadic closure acts within the IMDB. There are many potential variables, and it can be difficult to properly interpret all of them with a limited amount of computational power.

In short, while there is a good amount of evidence for triadic closure in its general sense, its use as a tool for gauging social networks is still questionable. More research needs to be done in order to find out how much of an impact triadic closure can have, as well as how it might be possible to improve the means of detecting triangles in the IMDB data set, and in social networks as a whole.

References

[1] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, MA: Cambridge University Press, 2010.

[2] S.A. Golder and S. Yardi, "Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality," in *2010 IEEE Second International Conference on Social Computing (SocialCom)*, 2010, pp.88-95, 20-22. doi: 10.1109/SocialCom.2010.22
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5590841&isnumber=5590391

[3] S. Garg, T. Gupta, N. Carlsson, and A. Mahanti. "Evolution of an online social aggregation network: an empirical study," in Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference (IMC '09), New York, NY, 2009, pp. 315-321.
doi: 10.1145/1644893.1644931
URL: http://doi.acm.org/10.1145/1644893.1644931

[4] S. Schnettler, "A structured overview of 50 years of small-world research," Yale University, New Haven, CT, 2009.
doi:10.1016/j.socnet.2008.12.004
URL:
http://www.sciencedirect.com/science/article/pii/S0378873308000695

[5] S. Dommers, R. V. D. Hofstad, and G. Hooghiemstra, "Diameters in preferential attachment models," 2010.
URL: http://www.win.tue.nl/~sdommers/files/DiamPA_DHH_rev.pdf

[6] B.W. Herr, W. Ke, E. Hardy, K. Borner , "Movies and Actors: Mapping the Internet Movie Database," in *Information Visualization, 2007. IV '07. 11th International Conference,* 2007, pp.465-469.
doi: 10.1109/IV.2007.78
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4272022&isnumber=4271944

[7] A. Esfahania, "Connectivity Algorithms." URL:
http://www.cse.msu.edu/~cse835/Papers/Graph_connectivity_revised.pdf

[8] *IMBD Actor-Movies Data* [Online], Available:
http://www.cs.rpi.edu/~magdon/courses/CASPdata/IMDB.zip

[9] *Jgraph*.[Online], Available: http://www.jgraph.com/

[10] E. N. Gilbert, "Random Graphs", Bell Telephone Laboratories, Inc., Murrah Hill, NJ. URL:
http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.aoms/1177706098

[11] M. S. Granovetter, "The Strength of Weak Ties," in *American Journal of Sociology*, vol. 4, no. 6, pp. 1360-1380, May, 1973.
URL: http://www.stanford.edu/dept/soc/people/mgranovetter/documents/granstrengthweakties.pdf