Fractalizing Social Networks

Silvia Cobialca, Juan M Ale

Universidad Austral, Buenos Aires, Argentina
silvia.cobialca@accenture.com, ale@acm.org

Abstract. Fractals are self-similar structures that exist widely in nature. We are aiming the current work to prove that social networks, although not a naturally generated structure but one created by humans within the World Wide Web, show a fractal behavior as well and as such, will experience a self-similar kind of evolution.

In the present work we attempt to find through the study of fractal behavior, how the introduction of a new element in the social network will impact in the existing network structure and in the network growth. Also our main interest is into how the new node will start interacting with the existing communities in order to eventually build its own.

Keywords: Social Networks, Fractals, fractal dimension, box dimension, adjacency matrix, fractal social models and algorithms.

1 Introduction

The main focus of this work is the idea of social networks as fractals and how to explore and elaborate from there in order to build new functionality useful to the study of their behavior over time. If social networks behave like fractals then, the existing network random generators would not apply to them as their randomness lose their meaning and we will need specific generators with proper parameters more suitable to be applied to fractal behavior.

We show how fractal nature applies to social network structures and how their evolution can eventually be predicted by modeling upon them. To do so, our starting point is to reviewthe basics about networks in general and their parameters, measures, types and the existing models used to generate network structures, to later focus on social networks in particular. We also review fractal theory and its applications to be able to merge the two concepts together while working on the models.

As a motivation, we noticed that current models for networks are based on randomness in a general way and sometimes such models donot take into account the nature of the network. In that sense, we consider that social interaction has a strong relevance that should be taken into account when studying models that will be used in the future for social networks. We base the present work on the basic structure of the social network and escalate from there in order to formulate the prediction as precisely as possible thinking more in the individual components and how their behavior will imprint its pattern in the whole network.

Our main contribution is a model and algorithms that allow us to show that evolution of a social network obeys fractal rules.

After reviewing previous work on this matter we have come with the ones more relevant to our line of thought which are detailed below:

Leskovec (2008) studied network evolution, network cascades and large data while analyzing large social networks as a whole in order to formalize and try to predict future behavior and structure. His idea was very nicely worked out by a three by three focus on observations, models and algorithms on network evolution, cascades and large data. He worked mostly on the network itself, but not from an individual node point of view.

Song et al. (2005) analyzed several real networks and found that they consist of self-repeating patterns on all length scales.

In the work by Faloustos et al. (2006),the authors also investigated network structures and expanded about the graph generators, but although they mentioned power laws they didn’t elaborate on the possibility of using fractal similarity to build the network synergy expressed by the links between nodes.

Erdös, P., Renyi, A.(1959) set the base for future contributors to the area by defining random network generation and probability of connections between nodes among other definitions.

We use a different approach from the previous work mentioned above as our main focus is the idea brought by Benoit Mandelbrot (1982). The author defined fractal structures and the fact that they were present in nature in several levels of complexity; he also mentioned that the interaction between systems can be seen as fractals. This is, in fact, the driver of our present work.

The remainder of this article is organized as follows. In Section2 we provide background information on the main topics of this work. In Section 3 we present our Fractal Social Network Model and our proposed algorithms. Some empirical results are shown and discussed in Section 4. Finally, we conclude our work in Section 5

2 Background information

Social networks.

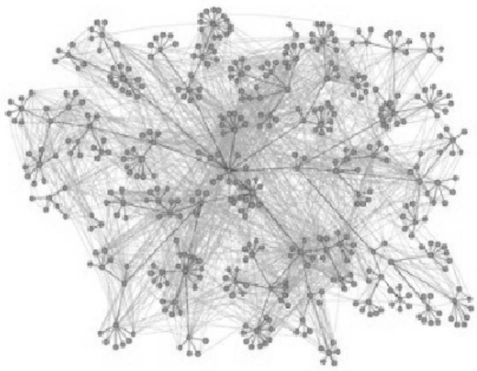
Actual social networks are more related to collections of social ties among friends, like the ones based on the internet as Facebook, Twitter, Instagram, Tumblr, Flickr and many others; although there are other networks focused on businesses or professional relationships like LinkedIn, WordPress, Yelp, etc. These examples are all among online networks which, because of their nature, can be massive and reach farther boundaries than those based on people direct interactions, which can be geographically based, although both kinds are good examples of social networks where the same theory can be applied.

Social interaction have grown steadily in complexity over the course of human history, due to technologicaladvances facilitating distant travel, global communication, and digital interaction as mentioned by Kadushin (2012).

These networks can be seen as graphs with nodes (the individuals participating in the network) connected by links as shown in Figure 1.

Understanding any one piece of information in this environment depends on understandingthe way it is endorsed by, and refers to, other pieces of information within a large network oflinks.

Fig.1.*A representation of a social network based on email communications*
(Image from http://www.personal.umich.edu/ladamic/img/hplabsemailhierarchy.jpg)



We represent social networks as graphs because a graph is a way of specifying relationships among a collectionof items. A graph consists of a set of objects, called nodes, with certain pairs of theseobjects connected by links called edges.

Graphs are defined as *directed* if they consist of a set of nodes together with a set of *directed edges* each directed edge is a link from one node to another, with the direction being important. Directed graphs are generally drawn as in with edges represented by arrows. When we want to emphasize that a graph is not directed, we can refer to it as an *undirected* graph.

Random graph algorithms

From the initial work done by Erdös and Renyi(1959) to the present, several algorithms have been developed in order to generate social network graphs (or graphs in general) and study their evolution over different epochs (a way to call the parameter to measure passing time). The simplest algorithm was one of complete randomness where the probability of two nodes connecting (or contacting) each other was the same for every pair of nodes belonging to the network.

Of course, this approach is too simple and no real life network will behave that way, more likely the nodes involved within the network will connect with other nodes based on preferences, similarities, recommendations from others but they hardly will connect in a random fashion.

There are also other kinds of social networks like the ones based on*small-world phenomenon* in which the applied logic states that any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances. This has been proved to be truth by several previous investigations being the oneconducted by Milgram(1967) the most popular, but we also have reviewed the work from Mathias and Gopal (2000), as theyall reveal that often we meet a stranger and discover that we have an acquaintance in common. Recent work has suggested that the phenomenon is also existent in networks arising in nature and technology, and a fundamental ingredient in the structural evolution of the World Wide Web. We explore if the "fractal" network as we call it, exhibits also a sort of small-world phenomenon in its behavior.

What are fractals?

First of all, and before getting to the point of a definition of a fractal let's take a moment to imagine what we currently denote as chaos, or chaotic behaviorusually related to some unpredicted pattern that cannot be formalized in any way,as studied by Shroeder (1991),. The difficulty of working with such behavior is, of course, the inability to adjust it to any existent law that could rule it and help to the job at hand.

Fractals are self-similarity structures that can explain this behavior and bring certainty and predictability whereas there was chaos and misinterpretation beforeThe trick is to understand the structure itself and how it evolves and grows from the basic initial unit.

Fractals are useful in modeling and explaining natural complex patterns that can't be explained by Euclidean geometry. In these irregular and fragmented patterns, we can see how nature expresses itself in leaves, mountains, turbulences and also inside of us in our blood vessels or pulmonary systems. They are all examples of shapes that can be built by scaling up a base structure over and over, which implies a certain degree of irregularity, but in an unusual regular way at all scales.Barnsley (1988) in his work compiled different fractals existent in nature like forests, mountains and landscapes in different parts of the world, and also reviewed the theory behind their existence.

To understand that, we need to find the fractal dimension and the basic structure that the fractal is built upon and later grow it from there.

Defining the fractal dimension.

Fractal Dimension.

The concept of fractals started to take form when Benoit Mandelbrot vocalized his idea of a continuous escalating structure found in nature in different organic and inorganic systems.

The fractal dimension is a measure that will indicate the relationship between the size of the individual smallest structure that comprises the fractal and the total size. The formula is as:

$$D=\log N(r)/\log (1/r) \quad (1)$$

Where

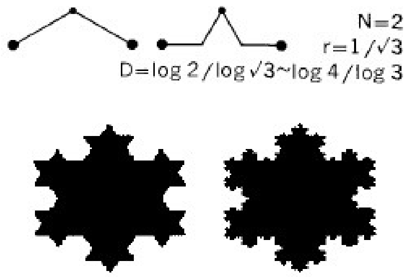
D: Fractal (or Hausdorff) Dimension

N: number of base parts

r: similarity ratio

There are a few examples that can help us illustrate this definition in a way that clearly enlightens our knowledge and understanding of it. One example is the Koch coastline shown in Figure 2.Every line the triangle is made of is added another pair of lines in the middle and the pattern is repeated over and over in every side finally getting the Koch triangle in black below the main structure.

Fig.2. Triad Koch coastline Fractal Dimension



Box Dimension.

There is another way to measure the fractal dimension and it is by sizing the smallest squared box that will include the base structure and counting the amount of those boxes that fit in the whole fractal structure. The number of boxes is N(r) where r is the size of the side of the square used as box dimension. Then we can calculate the fractal dimension using (1).

Multi-fractal Dimension.

As with the normal fractals defined above, there exists also a category of fractals that donot have one only fractal dimension and because of their nature, since they were built from, say, bricks of different sizes, they can be called multi-fractals.

Hence, the trick with multi-fractals is to identify the base “bricks” they are made of. Examples of multi-fractal structures are the diffusion-limited aggregations (DLA) like the ones generated by the colloids where the structure grows one molecule at a time. An image of a multi-fractal is shown below in Figure 3.

Fig.3. An example showing a multi-fractal

Are all social networks fractals?

After reviewing the information provided above, we can start to think that social networks seem to behave like multi-fractal structures. We cannot at this point of our investigation ascertain which would be the fractal dimension of them and more experimentation is needed for us to be able to provide such metrics.

For now, let's just hypothesize about how the behavior and the way they get generated seem to be compliant with that of multi-fractals.

3 The Fractal social network model

3.1 The model.

First of all, it has been observed during the initial investigations that the introduction of a new node v in the network is always through the knowledge of at least one pre-existent node w and as that happens, even if the new node has been introduced into the whole network, only a relative small region is available to it. Nobody expects the newcomer to start interacting with every community in the network right away, and for that interaction to start, time is of essence. We call that amount of time $T_0$the "introduction time" which is a grace period allowed to the new node before it gets the first contact with other nodes existent in a network and that is different than its sponsor in the immediate network (or community). The introduction time depends on the size of the community where node w exists, and of course the ties between v and w which are unknown and different to every case, so this is another parameter to take into account at the time of the on-boarding of the new node.

Once node v starts its interaction with the nodes in w's community, there will be only a certain amount of time T_1 in which it will start to reach out to nodes belonging to the neighborhood in a way that will be proportional to the amount of connections its newly acquired community has with the outside communities. This means that, if node w community C_0 has connections with three other communities C_1, C_2 and C_3 but the connectedness between them is, say n_1, n_2 and n_3 in which n_1>> n_2>> n_3 then the interactions between v and the nodes in said communities will start before with C_1 than with C_2 or C_3.

Now, the next item to take into consideration is the kind of node v will become within time, meaning if it will be a popular node or the opposite, more like a shy node. That will depend on the willingness it has to share and activate new connections with the rest of the network. In other words, if v is already a popular and well known individual when it gets in the network, there is a great possibility that it will "attract" the attention of the other nodes and they will reach out to it to connect and become popular as well. But if v is an individual that got in the network for a specific task and nothing else, it is improbable that its connections will increase beyond what is expected from it and so its degree of connectedness will be very small.

Fromthe adjacency matrix and the fractal dimension we get the minimum structure to be replicated in order to get what becomes the node evolution and future participation in the network. This can be done in several iterations and with different nodes in order to get the final base structures of the network.

We are talking about base structures because we consider the social network to be a multi-fractal and as such, it is built upon several box dimensions. We describe in detail this procedure in the next subsection.

There is a pre-condition to be taken into account in this algorithm, which is that a network should exist prior to the application of this algorithm. This way, the new incoming nodes can have the base layers of the existent network and they canreplicate similar pre-existent structures as theyactivate their new connections.This is key to ensure self-similarity patterns.

3.2 The fractal connection algorithms

We present the GetBox algorithm which will get the box dimension specific to a node in a certain community. It finds all the nodes the given node is likely to connect in order to keep the self-similarity structure in the network and its current connections.

```
Algorithm:GetBox(N,Adj[N,N],v,Cm[N])

Returns Bv list with nodes for potential connections

For each v in Adj[N,N]

Oudg[v]:=outdegreet(v,Cm(v))--the outdegree of v within its
communities of interest

End For

Freq:=Frequency(Oudg(v))   --we calculate the frequency
distribution of all nodes' outdegreeconnections

Freq=Rnd(Fr)      --we select randomly one of the existent
Frequencies in the network to be the box dimension for i at t+dt

Create Empty List(Bv)

i=v

for each j in Adjt[N,N]

If Adjt[i,j]=1or j in Cm[i] and Cardinality(Bv)<Freq--the
frequency is a measure of the box dimension, since there are several
boxes we use one of the available options randomly

Then Add j to Bv    --Bvwill contain all the nodes v should be
connected to in t+dt

End For

End.
```

The ApplyBox algorithm will use the Box dimension obtained by the GetBox algorithm in order to apply those new connections to the given node (and keep the existent ones)

Algorithm:*ApplyBox(v,Adj$_t$[N,N],w$_f$,B$_v$)*

```
We assume v has been in the network since time t

Begin

i:=v         --we search in row corresponding to v node

For each jin Adjt+dt[N,N]

If Adjt+dt[i,j]=0

ifj in Bv and wf(j)>0.6  --the higher the wf the more
willing to connect and the longer will be Bv list

Adjt+dt[i,j]=1

End For

End.
```

List of variables

Willingness Factor: w_f (used to distinguish nodes with interest of acquiring new connections from others not that interested in any interaction)

Size of the network: N (number of nodes in the network)

Introduction time: t (the time at which the new node is included in the network)

Adjacency matrix: Adj [N,N] (a matrix which describes a graph by representing which vertices are adjacent to which other vertices)

Cm array: community [N] (we don't know at first how many communities exist, at a maximum it can be the same as the amount of nodes)

Box dimension list: B_v(list with nodes that will likely get connected to the new node in order to preserve self-similarity in the network)

4 Experimentation: A simple example and findings

To experiment our theory we have a Facebook network of 484 nodes and 33272 links where we are able to see interactions between two timeframes.

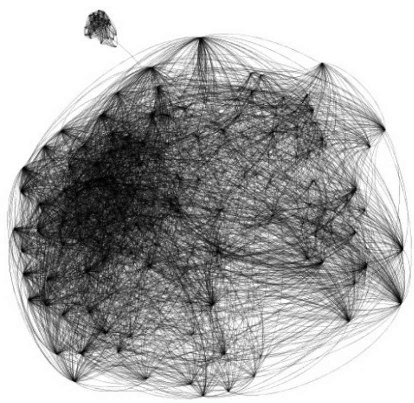
We collected the network and classified the nodes with more activity, the new nodes and the nodes that left the network between timeframes *t* and *t+dt.* The network is shown below in figure 4.

Below are listed the key findings during our experimentation:

- The willingness is a very important parameter to be considered in the existent nodes as well as in the newcomers. It has relevance in the connections growth because if the nodes are not interested in connecting with other existent nodes they will not attempt any new interactions, even if they have the best opportunities by being linked to the most popular nodes and communities.

- The frequency distribution of outbound connections (we are for now interested in the outbound) seems to be related to the different fractal dimensions of the boxes.
- There were several nodes observed at *t* with sponsors among the nodes with more connectedness and they still didn't increase their degree of connection at *t+dt*. We believe that these nodes weren't interested in new connections.
- One of the nodes with highest degree of connections didn't increase its connectedness in the new timeframe which contradicts the "rich get richer" principle. This behavior can also be explained by the willingness factor in a way of saying that there were no interesting things out there for this node to grow to.
- There are also some nodes that shrunk during the timeframe considered in our study, which is not a surprise as due to the network dynamics it's expected a certain level of change in both ways for the nodes degree of connections. This fact is still worth mentioned and something to be investigated and expanded in the future as well.
- The nodes with more growth grew beyond their community of origin and we can infer that they are then more mature in a way that they can reach and interact with new communities. It is expected that in the future they will continue to grow in this same way. The new communities were known from the most popular communities in the list of the pre-existent linked nodes. This is exactly what we are taking into consideration while developing theGetBox and ApplyBox algorithms.

Fig. 4. The network used in the experimentation phase (notice the new small community at the top left of the graph)



5 Conclusions and future work

This is a work in progress and in the following paragraphs we are presenting the first results of this research.

We have come to understand from our work during the experimentation phase and the theoretical background that some of the interactions in social networks can't be taken as random and more so, the people making the present social networks act some times in ways that seem to be mimicking other people behavior. Hence our self-similarity approach seems to be more suitable for them than randomness.

We have also uncovered the existence of several parameters to be taken into account when modeling social network.

We are leaving for future work and enhancements of the model, the task of testingthe algorithms presented in this work in a more global social network, also the formalization of rules to prevent starvation of network components in order to ensure that all the components are added to the fractal structure while the evolution happens as well as how we explain the dynamics of the shrinking patterns in the nodes connections.

6 References

1.Leskovec J.: Doctoral Thesis: Dynamics of large networks. CMU-ML-08-111 (2008)

2. FaloustosC.,ChakrabartiD.: Graph Mining: Laws, Generators and Tools. ACM Computing Surveys, vol 38 (2006)

3.Shroeder M.: Fractals, chaos and power laws. W. H. Freeman and Company (1991)

4. Mandelbrot, B.: The Fractal Geometry of Nature. International Machine Business (1977)

5. Barnsley M.: Fractals Everywhere. School of Mathematics. Georgia Institute of Technology (1988)

6. Mathias N., Gopal V.: Small Worlds: How and why. Department of Computer Science and Automation (2000)

7. Erdös, P., Renyi, A.: Of the Evolution of Random Graphs. (1959)

8. MilgramS., The small world problem. Psychology Today (1967)

9. Burt R.: Structural Holes and Good Ideas (2004)

10.Kadushin C.: Understanding Social Networks. Oxford University Press (2012)

11. Song, C., Havlin, S., Makae, H.: Self-similarity of Complex Networks. Nature, Vol 433. Nature Publishing Group (2005).