ΜΙΧΑΗΛ ΜΟΥΔΑΤΣΟΣ
AM 387

ΜΙΧΑΗΛ ΜΟΥΔΑΤΣΟΣ
AM 387

On natural mobility models

(1) SUMMARY

1. The authors observe that current mobility models are unrealistic at a time when mobility is more and more incorporated in networking. They will attempt to form a model from observations in real-life networks, focusing on their observed scale-free spatial distribution. They claim that their work could apply complementary, on various areas, from networking to sociology and biology, where large-scale measurements are difficult in practice.

2. The authors give us an overview an overview of the existing mobility models, which are constituted by the individual and group models. The current individual mobility models (in which each node moves individually) are the Random Walk model, where the nodes change direction and speed randomly, the Random Waypoint model, where nodes change destinations randomly, the Random Direction model, which differs from the "Waypoint" to the fact that nodes can pause at borders, the the Boundless Simulation Area model, where there are no bounds using a toroid, the Gauss-Markov model, which incorporates memory of previous state along with randomness and the City Section model, which uses a street-network map along with movement constraints. The current group mobility models (in which each node moves with respect to its group's movement) are the Reference Point model, where group of nodes move using one of the pre-mentioned models, the Exponential Correlated mobility model, which uses an exponential function to calculate the next group movement, the Nomadic Community model, where the group moves in a "Waypoint" style and each group node in a "Walk" style and the Pursue model, where the group follows a node having its own mobility model.

3. The authors then move on to the observation of the aspects of real-life networks. The two main aspects from which all common features of these networks originate, are the scale free property and the high clustering coefficient. In biology, an organism's metabolism form networks of substrates connected with reactions, the interactions between proteins form a network of protein bindings, the two dimensional polymers form conformation networks and the relation between gene expression form a network, all of which show either power-law degree distributions or high clustering coefficient or both, contrary to random graphs with Poisson degree distributions and low clustering coefficients. Again in computer networks the inter-router and inter-domain networks as well as the directed hyperlink networks show both power-law degree distributions and high and similar clustering coefficients. Finally, in sociology, the IMDB form a bipartite graph of movies and actors connected with participation having high clustering coefficient, citation networks in scientific publications form networks with power-law degree distributions and phone calls between phone numbers form networks with power-law degree distributions. Albert and Barabasi, in

their model reproduced the above phenomena using growth and preferential attachment.

4. Having shown the importance and regularity of the power law and relevant scale free phenomena, the authors remind the disadvantages of current group mobility models, being rigid and unrealistic, and express their intention to match reality using scale free distributions. Guided from people tendency to follow the mass in thoughts and acts and results from research on ad-hoc networks where scale-free inter contact distributions could also imply spatial scale free-distributions, they primarily present their notion of gathering mobility, which expresses the perpetual dislocation and reformation of different groups in life, because of occasional attractions. The authors then present the details of their conception whose basic elements are the individuals and the attractors, with the individuals moving from attractor to attractor moved by forces based on their distances from attractors and from other peoples choices with respect to attractors. Simulating their model for 10000 individuals resulted in a growth accompanied by a preferential attachment with scale free characteristics (e.g. the order of the attraction followed a power law).

5. The authors conclude their work citing that individuals in their model, although they move independently and don't belong explicitly to groups, they exhibit strong collective behaviour and are influenced by others, gathering around canters of interest of varying popularity levels. They state that scale-free-based growth of population resulted in scale-free spatial distributions and they placed the maintenance of this distribution in case of population decrease and renewal, as a future research goal.

(2) EVALUATION

In this paper the authors did a good use of crossing information from different research papers. They showed how inspired from a similar idea (by Jardosh et al.) but with a different approach, they resulted in a innovative and promissing result. Their model is amenable to extensions that could tune the effect of some of their ideas or that could add more but desired complexity.

(3) COMPARISON

As in the case of the router-level topology understanding paper, the authors use simulation methods to derive their conclusions. It's more a follow up work with respect to faloutsos' paper, which evades most, but not all, the deficiencies of power laws, described in router level topology paper, finding a suitable application in mobility patterns. One additional reason for which their work is not affected by the constraints presented in the other paper is that, apart from the difference in the research subject, there's also a difference in the applied technology, moving from wired to wireless networks, were the nodes now move and the edges and connectivity has different or even temporal meaning. Of course this does not mean that there aren't other constraints which could be applied to this case as well.

Comparing to faloutsos' paper they are slightly more descriptive with respect to the relation of preferential attachment and scale free terms with power laws, in the various aspects of life. In the other hand, the contributional part of their paper is quite small in terms of quantity and this forced them to make an extensive description of these terms as well as of all prior mobility models, which also aided them in proving their point. However, the paper of the router-level topology gives a better depiction of this phenomenon, while trying to render it unrealistic in its research field, by pointing out the process of the formation of the core highly connected cluster.

(4) CRITIQUE

Generally, the authors did not promissed many things and thus, were consistent with their intended goals. They did try to infer mobility from observations made in real life networks by forming the grouping mobility model. In this model they succeeded in mixing individual and group mobility with individual movement and destination choices constituted by forces proportional to other individuals' movement and destination choices. The only obscure part is how they resulted in the scale free spatial distribution, with respect to attractors' appearances in space. In other words the result in space is as dependent from the movement and destination choices of individuals, described by their model, as it is from the points were attractors appear in space and then disappear.

(5) SPECULATIONS, NEW QUESTIONS AND NEW IDEAS

One of the concerns one could express for this model is that it could fail to express long-period relationships, since it bears the danger of the creation of a giant cluster and regression from this cluster to another and back again to the first. As realistic as the authors might claim that this is, there's always a time in life, when the intrinsic attraction produced from the attractor to an individual diminishes, a phenomenon also called as boredom. This change of interest or need of refreshment could be added with a diminishing factor concerning an individual's attraction to an attractor. This diminishing factor could be a fraction, falling with respect to the time a node is constantly attracted by a specific attractor or is oscillating between this attractor and any other. This factor could rise up to the value of unity when the individual refrain from this attractor for "a long time". In the other hand one could say that this could be implicitly succeeded with the appearance and disappearance of attractors, yet the authors do not give any descriptions on the regularity of this phenomenon. In the latter case the fine-tuning of these appearances and disappearances would be the dual problem of simulating short and mid-period relationships more realistically.

On power law relationships of the internet topology

(1) SUMMARY

In this paper the goal is to prove the existence of power laws in internet topology. The authors initially prepare the ground by analyzing the basics of their work. Their analysis will be on both router and inter-domain level. They present degree of connectivity and path distances as the important metrics which describe graphs. With a small report on previous work they discuss some results such as the fact that two hops at router level represent one hop at inter-domain level, the generation of internet models requires simple structures for small networks or combinations for larger networks and they add that self similarity and heavy tailed distributions, which are related with power laws, are used to describe LAN and WAN traffic and power laws describe WWW traffic, pointing out that there hasn't been any work on power laws with respect to topology. Then, they present the data on which they performed their analysis. They used three inter-domain level instances of the internet which covered one year's time (97-98), in which the topology grew by 45%. They also used the router-level instance of the internet in 1995. Their analysis included the decomposition of the graph in tree and core components. The authors expressed that, conversely to statistics such as min, max and average values of metrics, which miss a lot of info and fail to describe skewed distributions, they want to characterize topologies with single numbers. Finally, they defined several graph metrics with which they will perform and evaluate their study and they denoted that they used the method of linear regression and the correlation coefficient to fit a plot to a line. The authors, then move on to their first power law: the rank exponent R. The outdegree of a node is proportional to the rank of this node in terms of outdegree raised to the power of the exponent R. The three inter-domain instances showed close exponent values (which in fact were the slopes of the plots in logx-logy scale) which shows how such a graph can be characterised from a single exponent value and its different value for the router-level graph implies that such exponents can also distinguish different families of graphs. The authors produce two lemmas which connect the rank exponent with each nodes degree (1st) and consequently with the graph's number of edges given the number of nodes (2nd). (This gave a 9%-20% error when applied in the real trace contrary to the 3,6% when using linear interpolation for the last instance, with the difference that the former method does not need previous instances). Then, follows the second power law: the outdegree exponent O. The frequency of an outdegree is proportional to the outdegree to a power of a constant O. The calculated exponent values have the same characteristics, with the exponent of the router-level graph to be very close to that of the inter-domain level possibly revealing a fundamental property of these networks. As such, the authors propose that we can test the realism of a graph, by testing if its metric follows a power law and if its exponent is close to realistic numbers. Subsequently, the authors define an approximation: the hop-plot exponent. The total number of pairs of nodes within any path of h hops, is proportional to that number of hops to the power of a constant H. They produce the third lemma which gives the formula of this approximation. Similarly with the rank exponent, the inter-domain level graph instances give very close exponent value, whereas the router-level gives a different value, being a sparser graph, which shows that the exponent H can distinguish different families of graphs. They also define the metric δ_{eff} (effective diameter) as a diameter of the graph which is calculated based on the hop plot exponent and the number of nodes, since lemma 2 gives the number of edges. They propose to use δ_{eff} as the number of hops a broadcast should travel in order to have a high probability to reach a desired destination and not to flood the network as well. They show that on real data δ_{eff} covers 80% of the total node pairs within δ_{eff} hops. Using the hop plot formula they come to lemma4 with which we can calculate the average size of the neighborhood within h hops. They prove that lemma4 produces results closer to reality than the formula with which we calculated the average neighborhood size so far and which was based on the average outdegree of the nodes in the neighborhood. They also propose the model of an H-dimensional sphere with radius the number of hops, over the previously implied as exponential in the number of hops model, concerning the h-hop neighborhood and its size. The third power law is then presented: the eigen exponent ε. The eigen value λ of a graph is related with the graph's adjacency matrix A ($Ax = \lambda x$) and properties such as diameter, the number of edges, the number of spanning trees, the number of CCs and the number of walks of a certain length between vertices. The eigenvalues λ_i of a graph are proportional to the order, i, to the power of a constant, ε. Similarly with the rank exponent, the three inter-domain level graph instances have almost the same exponent value, showing again a property insensitive to graph change in size, whereas the router level graph has a different value, which implies that the eigen exponent can distinguish different families of graphs. In the last part of the paper the authors give the practical uses of the power laws, the scope of the deriving predictions and the intuition behind the existence of power laws: The exponents describe the skewed distributions better than averages which imply uniform distributions. The calculated metrics can contribute to protocol design and performance with an improved $O(d \cdot h^H)$ estimate of the average neighborhood size, over the, until now, $O(d^h)$. Given a number of nodes hypothesis, prediction can be made for the number of edges and the effective diameter. Since coincidence is excluded from the high correlation coefficients, the intuition behind power laws lies in the fact that a small change in the topology is not autonomous and affects gradually, like a fading wave, the rest of topology and thus, maintaining its fundamental characteristics. Moreover many other systems and networks of different natures are governed by power laws, from biological or automobile networks to web traffic and income distributions. Finally, they sum up their work pointing out their contributions. The three power laws, how they connect with useful graph metrics and how they can be used to predict such metrics and to assess the reality of simulated topologies.

(2) EVALUATION

What was quite interesting was the explanation of power-law 1, which in short expresses the trade-offs in adding another edge to a node with high degree versus a node with low degree both from a financial and functional point of view. Moreover its very interesting that power laws describe many other networks, even biological ones. As far as their methodology is concerned, they avoid using averages because, indeed, they can be misleading, since a few very high values can alter the average value so much that it won't express the majority of the values. In fact this is exactly what happens in this case with a very small number of nodes with high values of outdegree, which alter the mean so much that 85% of the nodes have a lower outdegree. On the contrary the exponents succeed to describe the whole plot since it's its slope in log-log scale. The authors performed their experiments using real data. In the case of the inter-domain level graphs they used more than one data traces. In this way they can test with more confidence the validity of their hypotheses. Moreover they test the resulting lemmas of their laws, on the data in order to verify their accuracy. They also check the sensitivity of their Rank exponent lemma, with respect to the exponent value and the resulting number of edges. In this way, one can further estimate how the resulting values could vary with respect to a possible error on the Rank exponent used.

(3) COMPARISON

Since there was not in fact any other work on relating the internet topology and its characteristics with power laws, this paper can be considered an addition to the up-till-now difficult work to model the internet. The paper gives metrics which fit the methods the latest simulators create topologies and improves methods which can be used in protocols and protocol analyses, giving a novel and faster method to estimate the average neighborhood size with cost O(d*h^H) over previous O(d^h) and proposing the number of pairs within h hops, as a metric of the density of a graph.

(4) CRITIQUE

Apparently, the authors cover practically all the issues that they had put on the table (*). The only obscure part is how their work connected with the fact that they chose to decompose the graphs into tree and core components, which is, as they state, the state-of-the-art model. Apart from this point, the paper is generally very well constituted, enumerates clearly its goals and results but also points out its weaknesses and where the results should be taken under consideration with caution. For example, they state in the introduction that the sample space is limited to make any generalizations and especially in the case of the hop plot exponent they admit that the, by-nature, small number of points in the plots restrain them from defining another power law, resulting in an approximation.

(*) (They derive estimates of topological parameters (average number of neighbors within h hops). They do identify three power laws for the topology of the internet. They do show that there is some exponent for each graph instance. Indeed, they introduce a graph metric to quantify the density of the graph, the P(h). They do show how to use their power laws and their approximation to estimate useful parameters of the internet - the number of pairs within h hops P(h), the effective diameter δ_{eff}, the number of edges E and the average neighborhood size within h hops NN(h). They do measure several crucial parameters for the most recent graph generator - the δ_{eff} and the number of edges E.)

(5) SPECULATIONS, NEW QUESTIONS AND NEW IDEAS

There were some issues in the paper though, who put us under some consideration.

In the case of the outdegree exponent O they have excluded some very large outdegrees with frequency 1. Although this shows that such outdegrees occurred only once in the trace this does not reduce the significance of their existence, since they exist and they represent nodes which connect to many others as consequence of a physical phenomenon which needs further investigation than ignoring it. The existence of a such small percentage with high outdegree implies that the rank exponent possibly describes a large part of reality but not totally, as well as the existence of a tail in this 'claimed' linear nature (in log scale) of such metrics; hence when we try to simulate such topologies we should apply additional info for a more realistic simulation, if we don't want to miss such hot spots. In the other hand they claim that the rank exponent captures the nodes with very high outdegree

Generally, their observations are "a posteriori" conclusions. When the value of an exponent is different for the inter-domain level graph than that of the router level graph they claim that this exponent can distinguish different families of graphs. In the other hand, when the exponent value was very close in both cases they claimed that this exponent captured a fundamental property of the network. This flexibility is convenient but also feels a bit unstable.

In the case of the effective diameter they claim that they can find the number of hops within we can reach nearly 80% of all the nodes and the only thing we need to know is N and the hop-plot exponent H. In addition they state that the number of edges can be computed using lemma 2. However does the latter argument take into account the fact that the lemma 2 computes the number of edges with 9 - 20 % error? It would be interesting to see whether this affects the results in a way that less than 80% of the nodes are reachable within a path of δ_{eff} hops.

In the case of the average size of the neighborhood the authors do not comment the fact that their approach both underestimates and overestimates the real values. Although their estimates are closer to the real values, than those of the average degree approach, they do not show how their formula works for neighborhood sizes of more than 4 hops. They also do not state the reason they choose to stop at 4 hops. One could say that this was the number of the calculated effective diameter which again was one of their own results. The plot of their formula seems to ascend rapidly after 4 hops which might give a major overestimation of the average neighborhood size if the δ_{eff} is 5 after all. This overestimation could be comparable to the underestimation of the average-degree approach. Whether one of the two is preferred depends on the reason for which we need the average neighborhood size.

Finally it would be good if they could verify their router level assertions, with more than one instances in a follow-up workshop.

A First Principles Approach to Understanding the Internet's Router-level Topology

(1) SUMMARY

This paper analyzes the router level topology giving another more practical point of view. The authors first shortly present their work. They note that there's currently a tension to evaluate topologies using large-scale statistics as metrics. They comment this approach as incomplete since graphs with same node degree distributions can result from opposite graphs in terms of network engineering. From this viewpoint they use here the notion of "first-principles approach" to describe an attempt at identifying some minimal functional requirements and physical constraints needed to develop simple models of the Internet's router-level topology. They state that they will focus on a few critical technological and economic considerations that they claim provide insight into the types of network topologies that are possible, introducing the notions of network performance and network likelihood as a new means for discerning important differences between generated and real network topologies.

The authors then reference the evolution of previous work, from random graphs, to hierarchical structural models and then to degree-based topology generators. The latter has three probabilistic generation methods, the preferential attachment were a sequential addition of nodes is performed and, the general model of random graphs (GRG) and Power Law Random Graph (PLRG) which connect nodes under a a given (power law) degree sequence. They all tend to have a few centrally located and highly connected "hubs" through which essentially most traffic must flow, which is the essence of the so-called scale-free networks. The authors claim with the latest models' absence of an understanding of the drivers of network deployment and growth, it is difficult to identify the causal forces affecting large-scale network properties and even more difficult to predict future trends in network evolution.

As solution, the authors contradict their "first principles approach". They explore some of the practical constraints and tradeoffs at work in the construction of real networks. As technology constraints they introduce the feasible region which expresses the trade-off between connectivity and speed and also defines the upper limit (the frontier) of this region. Concerning the economic considerations, the author show that despite the possibility of tuning these trade-offs, the choice is affected by the tendency of the customers. End users usually pay for low speed service resulting in higher degree edge nodes, with high variability in the degree, in order to serve the many end users, whereas there's a traffic aggregation while moving to the core nodes, resulting in high speed routers with small degree. They conclude in "heuristically optimal networks": the previous considerations lead to heuristically optimal designs were a sparse mesh-like core is followed by a hierarchical tree-like structure at the edges, with the addition of high-speed low degree border nodes. The authors present two real internet back-bones, Abilene and CENIC, as a representative case.

As far as the evaluation of a topology is concerned, the authors claim that the metrics used so far to evaluate topologies such as, "node degree distribution", which can describe different graphs since we go from one graph to another using the degree-preserving rewiring technique, "expansion", "resilience" and "distortion", which give the same quantitative values for qualitatively different topologies and "hierarchy" as expressed by degree-based models, are inadequate and lack a direct networking interpretation. They define some performance related metrics, mainly throughput and router utilization, and some likelihood-related metrics, in order to differentiate between graphs of the same degree distribution.

The authors then move on to comparing simulated topologies with power law degree distributions but with different features, to prove their claims. As first example the authors compare topologies resulting from preferential attachment (PA), the GRG method with given expected node degree sequence, a generic heuristic optimal design, an Abilene-inspired heuristic design and a heuristic sub-optimal design, all of them having the same power law degree distribution. The results showed that degree-based generated graphs which had by and large, high likelihood showed extremely low performance due to their hub-like core which acts as a bottle neck for the whole topology. In the other hand the optimally designed topology had a low likelihood, being rare from a probabilistic graph point of view, but with careful bandwidth aggregation, they achieved high performance. Finally, low likelihood does not imply high performance with the sub-optimal design to have low, both metrics. In their second example they compare three topologies with the same core but different end user demand in bandwidth showing how this affects the degree distribution.

Concluding, the authors discuss the fact that different graphs, with average likelihood generated by degree-based models, are difficult to be distinguished with macroscopic statistic metrics. Moreover when, performance is involved these designs yield low performance where simple heuristically design topologies have high performance and efficiency. Then, the authors add that they haven't involved robustness in their analysis since it would complicate it, further than they would like in order to make their points, although it can be incorporated in their model. Finally, they defend their methodology which includes toy examples, denoting that they are sufficient for introducing their new ideas, for pointing out the need of incorporating them in present models and for comparison with them, and stating that validating their work on real data would be a goal for future work, since its a research topic itself. They end up pointing out the orthogonal nature of their paper with respect to current models, calling it a methodology paper which opens new lines of research.

(2) EVALUATION

One interesting point in the techical issues mentioned in the paper was the method of degree-preserving rewiring. A method with which you can transform a graph with a given degree distribution by changing it edges, resulting in a different graph with the same degree distribution. This was one of the key techniques used to build their arguments upon. What was contradictory and at the same time interesting is that the authors leave from the theoretical analysis of the problem, bringing up for discussion issues directly inspired from the practical view of the network engineering, yet without using real data, and they succeed in making their points. Consequently, they showed how even toy examples, such as the fictional graph representations they use, are quite useful to contradict directly different ideas, focusing on the desired aspect.

(3) COMPARISON

The authors in general differentiate from previous random or degree-based models. Conversely to faloutsos work, the authors focus on specific and practical characteristics which affect the structure of a topology, rather on graph theory terms and characteristics as conclusions of absolute statistical processing. Thus, their explanations are far more specific focusing on the exact practical technical and economical deployment characteristics, namely the available technology, the choice of the technology with respect to its functional position and the choice of the technology with respect to the end user demand, rather than philosophical speculations which can only try to formulate the problem in all its pre-mentioned dimensions. Furthermore, their work mainly has to do with the router-level whereas most of the faloutsos' trace comes from the AS level. However, both works claim that their findings apply (faloutsos') or are expected to apply (present work) on other network levels as well. In the other hand, the only real data the authors use are two topologies which match their ideas about how to properly design a topology. They produce optimize topologies inspired from the real ones in order to contrast with the generated ones from the degree-based models, whereas faloutsoi work on real data gathered across one year's time, since its a measurements work.

One of their representative comments worth mentioning are that "there is a long-standing but little-known argument originally due to Mandelbrot which says in short that power law type distributions should be expected to arise ubiquitously for purely mathematical and statistical reasons and hence require no special explanation".

(4) CRITIQUE

There were several goals implicitly presented in the beginning of the paper, all of which were achieved (*). In the process of presenting their work through the paper other issues aroused, some of them were not fully explained or left vaguely, but intentionally, uncommented. For example at the definition of likelihood the authors state "defining the likelihood of a graph g (- (belongs to) G(ù) as the logarithm of the probability of that graph, conditioned on the actual degree sequence being equal to the expected degree sequence ù", without giving any formula. More over they add that the latter can be shown to be proportional to L(g), without again giving any formal representation. They end by explaining that for the purpose of this paper, they simply use the l(g) metric to differentiate between networks having one and the same degree distribution, and a detailed account of how this metric relates to notions such as graph self-similarity, likelihood, assortativity, and "scale-free" will appear elsewhere. However "elsewhere" was not defined explicitly as elsewhere in this paper or in a forthcoming one. The authors are sincere enough to professionally comment that one important feature of network design that has not been addressed is robustness of the network to the failure of nodes or links, when they could avoid mentioning it at all. However, they sufficiently comment that it would complicate their analysis further than its goal need to, which goal was to point out the incompleteness of current topology models.

(*) [The authors implied they'd:

-show that power laws consist an incomplete description since random graphs could result in given distributions with no network intrinsic meaning.

Done: result in hub-like structures in the core of the graph, which form unrealistic, low-performance core topologies.

-develop a basic understanding of the observed high variability in topology-related measurements and reconcile them with the reality of engineering design.

Done: Described how routers in core would be differently set up (low degree with high bandwidth) from routers at the edge (high degree to support many low-bandwidth end users). Also different user demands at the edge could result in variable degree.

-use here the notion of "first-principles approach" to describe an attempt at identifying some minimal functional requirements and physical constraints needed to develop simple models of the Internet's router-level topology that are at the same time illustrative, representative, insightful, and consistent with engineering reality.

Done: same as above.

-focus here on a few critical technological and economic considerations that they claim provide insight into the types of network topologies that are possible.

Done: they explain the technical and economical constraints

-provide evidence that network models of router-level connectivity whose construction is constrained by macroscopic statistics but is otherwise governed by randomness are inherently flawed.

Done: they show the relation of their high likelihood l(g), with the low yielding performance due to the bottle neck in the core of the topology.

-introduce the notions of network performance and network likelihood as a new means for discerning important differences between generated and real network topologies.

Done.

-show that incorporating fundamental design details is crucial to the understanding and evaluation of Internet topology.

Done.

]

(5) SPECULATIONS, NEW QUESTIONS AND NEW IDEAS

The first thing one would ask about this work is, does it apply to Autonomous System-level as well? The authors conclude by speculating that similar technological or economical considirations could exist at his level as well. Subsequently, one would like to see, possibly in a forthcoming work, how the authors' arguments comply with real data. Do real networks comply with their first principle approach and derived metrics? Finally it would be useful to publicitate a technical report where to describe thoroughly the technical issues that were left unexplained in the paper, mentioned in the previous section, for eaxample the exact prior definition of likelihood and how it's proportional to their L(g) and relative likelihood l(g).