PIMS CRG:
Algorithmic Theory of Networks

CRG Leaders: Petra Berenbrink, Funda Ergun (Simon Fraser University), Valerie King (University of Victoria)

Proposed start date: May 2012

1 Summary

The technology revolution of the 1990s and the 2000s owes much of its existence to the advances in computer networking technologies. These advances have made profound changes in how we model, construct/modify, maintain, use, and, ultimately, view our networks. In this document we propose to form a Collaborative Research Group to work on the theoretical foundations for new generation networks. That is, we will develop and analyze models, design and prove the correctness and optimality of algorithms. We will explore the theoretical limits of communication and computation in this new context.

2 Rationale and Objectives

As networking technologies evolve, our understanding of computer (and other types of) networks change in terms of their scale, structure, and functionality. Now, we, as the scientific and technical community, are faced with new networking paradigms which are complex, heterogenous, and data-intensive. The resulting theoretical and practical problems are multifaceted, requiring a wide range of expertise from many fields. In this proposal we bring together researchers from four major universities in Western Canada, with what we believe to be the necessary range of interests and expertise in different aspects of these networks and with the technical skills to investigate these problems. To this end, we propose to initiate a collaboration among Simon Fraser University, University of Victoria, University of Calgary, and University of British Columbia through a Collaborative Research Group within PIMS. The collaboration is to be on the modelling and algorithmic aspects of networks with special emphasis on sensor networks and probabilistic techniques.

Right now networking theory is not a fully coherent field; much of the related research is scattered across computer science, mathematics, and engineering, mostly based on the approach used, or the particular application. For instance, publications pertaining to sensor networks can be found in venues dedicated to approximation algorithms, graph theory, information theory, distributed systems, randomized algorithms, data mining, computational geometry, or game theory, where their conceptual overlap with, more abstract topics such as metric embeddings, is lost. Theoretical networking research in Western Canada shares this pattern; there are several independent algorithmic groups in the region, generating significant activity in fields such as network data dissemination, distributed computing, decentralized coordination, data aggregation, coverage, power efficiency, and modelling. These groups share significant elements in terms of the general theme (distributed decision-making, sensor networks, game theoretic questions) while exhibiting complementary technical expertise. So far they have been working and publishing without much coordination or collaboration. We aim at bringing the expertise of these researchers together to improve our understanding of the new types of networks that have become so important for our lives.

The following scientific and technological challenges provide motivation for this work:

- *Increased quantities of data:* Modern user and system applications generate and store previously unthinkable amounts and data. For example, the amount of data stored by Yahoo, waiting to be analyzed, is measured in *peta*bytes. Cisco projects that the Internet traffic will amount to 1 *zetta*byte (10^{21}) bytes) over 2015.

  With such unthinkably large numbers, traditional notions of "feasible computation", as being computations which can be done polynomial time and space, are meaningless. The basic assumption that data can be held in memory and accessed at will does not hold any more. Instead, we may have sequential access to the data, have the data distributed over a large number of network nodes, or have access to random snippets of the data, like in the case with genome sequencing. Furthermore, even processing time that is linear or slightly superlinear in the input size can become overwhelming. As a result, some compromise has to be made, and typically accuracy will be the parameter which will be sacrified. This will result in interesting resource/precision tradeoffs.

  Along with the size of our data, our models are changing. Data can be a real time generated flow of ever-changing input, possibly with noise or unknown parts, as with the web graph. There may be simultaneous updates to the data from various sites. Our existing models are not satisfactorily capable of handling, or even representing these new situations.

- *New constraints on memory and power:* Data transmission is now very easy to achieve, even on the smallest and cheapest of devices. This has led to non-traditional types of networks with new kinds of constraints. For instance, sensor networks are networks in 2- or 3-dimensions in which the components are presumed to have very small memories and short battery life; the ability to communicate with one another is determined primarily by physical proximity. Low energy consumption used to be the domain of electrical engineers. Now theoretical computer science and mathematics now have to deal with such technical restrictions in order to develop models and efficient algorithms for these networks. These restrictions can be translated into graph theory, combinatorics, or discrete optimisation, and the behaviour of the nodes can be modelled as selfish agent and then analysed in game theoretic terms.

- *Heterogeneity:* The act of joining the Internet is a matter of using the TCP/IP protocol stack at a gateway router; as a result, it is quite easy to connect many different kinds of networks, such as the phone network, sensor networks, wireless and/or satellite networks, ad hoc networks, etc, to the Internet. This results in a structure which is very heterogenous and unpredictable. The nodes of today's network may be agents which are malicious, e.g. botnets, or simply users with varying objectives. The area of distributed computing has historically dealt with the coordination of nodes in small networks where identities of the nodes are known to all and faults are typically crash failures. Algorithms have involved either shared memory or all-to-all message passing. The challenge now is to get beyond these assumptions since they are infeasible for very large networks.

- *Social nature of networks:* We have now entered an era in which implicit social and economic relations have become explicit and open to study, while vast new kinds of social relationships have been created. What was before a circle of friends is now an explicit set of links in cyberspace. What used to be common wisdom shared by members of a group is now shared with hundreds of friends through Facebook. With these new definitions, we have new structures and new problems. The mathematical understanding of social networks, their structure, the spread of information and influence, has barely begun and will impact the study of sociology and other social sciences.

  There is increasingly more information about other naturally occurring communication networks such as biological networks, ant colonies, and bird flocking. Modelling collective behaviour in order to determining what these networks are capable of computing remains an important challenge.

- *New parallel programming paradigms* such as map-reduce or hadoop have been developed for massively parallel systems (involving hundreds of thousands of processors), which are now in use at Google and available for hire (cloud computing) at, for instance, Amazon. These systems are not adequately modelled by shared memory models nor by the fixed topology networks which were previously studied. They require a design which which allows for efficient data partition that minimizes communication costs between processors as well as minimising the rounds of computation. Because of the very large number of involved components, the algorithm design must be robust to faulty components and the systems have to be able to quickly recover from a faulty computation. A graph-theoretic (or game theoretic) understanding of these systems is currently lacking. More theoretical investigation is needed to better understand optimisation of computation in this area.

2.1 Research Directions

The teams in our participating institutions share interests but exhibit different viewpoints. Our SFU team has expertise on energy efficiency and data aggregation in sensor networks, random walks, sampling, massive data management, network modelling, routing in networks, game theoretic approaches for networks, and load balancing in distributed systems. The UVic team focuses on decentralised robust coordination, energy efficiency in sensor networks, game theoretic approaches to agent behaviour in networks, complexity, modelling of social networks, and foundations of communication and complexity. The University of Calgary team focuses on distributed decision making such as leader election in a network and communication in social networks, as well as multiprocessor network algorithms. The UBC team considers network modelling, routing and coverage in sensor networks from a computational geometry view, as well as from an discrete optimisation, information theory, and complexity theory perspective. All teams use, among other methods, randomised algorithms and probabilistic techniques in their work.

With the above strengths in mind, we propose several closely related research directions and present a few examples of the kinds of problems which lend themselves well to the variation in the skills of our participants.

Network generation models (Berenbrink, Kapron, King, Kirkpatrick, Sahinalp, Srinivasan, Woelfel) The graphs underlying certain naturally occurring networks such as the Internet and the Web Graph (defined by web pages and http links connecting them)exhibit distinct connectivity-related properties regarding their degree distributions, subgraph structures, etc. We are interested in developing randomized iterative models with compact descriptions that generate such graphs inductively. Our goal is to mathematically analyze the structure of such complex and changing networks in order to eventually understand how they came into being and how they have changed/will change over time. For instance, we can model the initial state of the network as a small clique, and at each step iterative add a new node v to the existing graph. The edges incident on v are defined in terms of the potential neighbours of v, selected from the existing vertex set according to some distribution dependent on the degrees of the vertices. In this model high degree vertices are favoured by the newly added nodes, resulting in a highly skewed degree distribution, generating *power law* graphs (see [1, 3, 2]). These graphs can be used to model social networks, the web graph, biological networks or the Internet. An added benefit from this work is network design, where, if we want to construct a network with certain properties, we would have a mathematically sound way of doing this.

In order to develop realistic models of complex networks one needs notions of similarity showing how close the models adhere to the real graphs that we observe. For this, we will define distance models in static (i.e., comparing two graphs) and dynamic (i.e., comparing the growth patterns of two graphs) settings in ways that capture their real world functionality. We will then use these models to improve our graph models, which, ultimately, sufficiently resemble their real-life counterparts sufficiently to allow predictions (for instance, with ϵ-approximations) of their structural properties.

One specific direction which we will explore is the generation of power law graphs through copy operations, as well as explaining the existence of higher order structures in social networks going beyond the degree distribution (see [4]). We will investigate the implications of these findings in biological networks as well. To analyze graph generation models there are two main methods that we will employ. The first one uses simple Martingale techniques to analyze the structural properties (such as degree sequences, number of triangles) of the graphs. The second one models the change of the network over time (for instance the change in the degree distribution) with sets of differential equations and then looks for solutions of these systems.

Randomized algorithms for network communication (Berenbrink, Ergun, King, Harvey, Higham, Srinivasan, Woelfel) In this sub-area we propose to work mainly in the areas of fault tolerant communication and randomized message passing/data distribution in networks.

In the area of fault-tolerant communication, we are given a graph G with n nodes and a function f on n inputs. Here the nodes represent processors and the edges communication channels. Node i carries local input x_i and the goal is for each node to compute a function $f(x_1, \ldots, x_n)$ of all of the local inputs correctly using a limited amount of communication. A subset of the nodes may be faulty or malicious, and act in a way deviating from their prescribed algorithmic behavior, resulting in an incorrect computation of f. We will develop ways of making it hard for the malicious nodes to influence the computation, and provide lower bounds on their numbers as well as the amount of communication that they need to perform in order to change the outcome. In general, randomizing the steps in the computation of f makes it impossible for the malicious agents to predict the next step and sabotage it. Using techniques from coding theory, one can make the function f less sensitive to corrupted inputs.

Our goal is to develop and mathematically analyze randomized algorithms for computing various general functions (such as represented in leader election, where all nodes agree on one node as the leader of the group, or byzantine agreement, where all nodes agree on a bit of information). We will prove upper and lower bounds for the multiparty communication complexity as well as for the running time in such settings. We will also develop an understanding of the tradeoff between malicious nodes vs. the number of bits of communication.

To analyze these results, we will use the Probabilistic Method of Erdos to show the existence of a clustering of nodes so that almost all clusters contain a majority of good nodes, regardless of the distribution of malicious nodes, as long as their number is bounded. Additionally, we can look at the problem from a game theoretic perspective where the malicious nodes are selfish agents, and predict the quality of the Nash equilibria we can expect to reach given their behaviour.

Our team will also investigate the related problem of randomized broadcasting and gossiping in networks. Here the randomisation makes the communication protocols robust against non-malicious node failure. To analyze broadcasting and gossiping in random networks one usually has to show many connectivity properties of random graphs (such as expansion), using algebraic tools and combinatorial techniques. One can then view message passing as similar to a random walk and analyse its properties given the expansion of the underlying graph. We will use various probabilistic tools to analyse these communication protocols, like tail estimates and Martingale techniques. In the case where gossiping fails to deliver the whole information, we will look into what kinds of guarantees can be obtained by sampling. Additionally, game-theoretic models of message-forwarding behaviour can be used to understand the demand that each node as a selfish agent is making on the capacities of its outgoing edges; we will investigate the outcomes of various strategies for message handling in the final outcome. Here we would like to focus on power law and social networks introduced in the last section and also other random graph models.

Another notion of fault tolerance is the real-life problem of keeping a network alive in the presence of link failures. In a more abstract sense, given a graph, we would like to assign edge-disjoint (possibly shortest) directed paths between every vertex pair. We would then like to *protect* each edge used against failures by assigning a secondary path to it, consisting of edges which are not in use for any path. When an edge on a path P fails, it is immediately replaced by its backup path, keeping P intact. However, the interactions between backup paths are complex, and the failure model (when an edge fails, it fails in both directions; when it is replaced, it can be replaced in one direction only) due to the real life infrastructure make the problem difficult. Our preliminary results show that k-edge connectivity in the original graph will not suffice for our network to remain connected in the presence of $k-1$ link failures. We will investigate how to use our resources most efficiently in order to maximize the number of failures that our network can handle, and explore upper and lower bounds.

Large data sets with restricted resources. (Berenbrink, Ergun, Harvey, King, Sahinalp) In Section 2 we discussed the increasing amount of data that has to be stored in networks. These massive data sets tend to have low entropy and they often contain strong structural properties. For instance, data can contain long monotone subsequences, periods, or large, often-repeated sections. Our goal is to extract information from this data such as its patterns or its dominant structural characteristics. We propose to devise methods to achieve this even when the data values are distributed across a number of locations.

Since the data sets are typically too large to fit into a processor's main memory *streaming* techniques have been developed to process the data by passing a window over the data and by processing the data in that window. The size of the window is upper bounded by that of the working memory. Usually only one sequential pass over the whole data set is allowed. In order to be able to remember the information content of the stream which has been read so far the data has to be compressed into a small *sketch*. This is simply a dimensionality reduction scheme which maps a vector of size n into a much smaller vector through some transformation. If our data set is a vector $x = (x_1, \ldots, x_n)$ of size n, a general way of sketching x is to multiply it by an $n \times m$ matrix A with values representing an algebraic object (such as a collection of univariate or multivariate polynomials) where the coefficients are randomly drawn according to some distribution. Typically, m is much smaller than n, resulting in a small sketch with some loss of information. If x is a k-sparse vector, obtaining (an approximation to) x from Ax is the problem of *sparse recovery*, closely related to that of *compressed sensing*. For our purposes of determining trends, typically polylogarithmic size sketches suffice.

We will develop techniques for calculating efficient sketches which preserve an approximation to the original information that we need so that we can eventually compute a function of the entire stream. While there is significant knowledge about sketching data for statistical functions, little is known about functions that depend on the order of the data (such as periodicity). We propose to use techniques from information theory, Fourier transforms, wavelets, and probabilistic techniques such as p-stable distributions and variance analysis, to extract information from distributed data presented as a stream. A quick look at some of the functions to be computed reveals that, in order to compute them, we may need to solve subproblems such as Hamiltonian Path or Traveling Salesman, which are known to be NP-hard. Quite often our problems are only special cases of these intractable problems that admit some (usually a constant factor) approximation. We also propose to show lower bounds for our models. These bounds typically use information theoretic techniques, some kind of fooling set for a particular VC dimension/direct-sums approach, and, most importantly, reductions from communication complexity problems. Properties of a more statistical nature can usually be lower bounded by reductions from two-player communication protocols. Problems that involve the ordering of the elements can typically be formulated as matrix problems which necessitates looking into multiplayer protocols.

Sensor networks (Ergun, Evans, King, Kirkpatrick, Harvey) A sensor is a small, wireless device with limited CPU power and memory, equipped with a small battery which allows it to perform a limited amount of communication over its lifetime. A sensor network can be modelled by a graph. The nodes of the graph are the sensors; an edge is assumed to exist between two nodes if their Euclidean distance is below a threshold d so that they can communicate cheaply. Each sensor v in the network generates a stream of data values V_i. The goal is to calculate at various points in time a global *aggregate* function $f(V)$ of the data values for all nodes at *one particular node*, called the root. Since the lifetime of the battery of a node is limited the amount of communication has to be minimized. A simple example for aggregate f is the average over all data values of each sensor, which is trivial to solve. Another example which sounds very similar is the median of the same, which is not trivial, and its approximation is the subject of many papers. Aggregation with limited communication using the supporting graph structures will be our main focus in this area.

Typically f is computed by calculating an *aggregation tree* as a subgraph of G. The data is then propagated up the tree, compressed at each internal node, and is finally collected at the root. Efficient techniques for constructing these trees are well known and are not the focus of this proposal. We instead propose to design techniques of computing f that minimize the amount of data that has to be sent to the root, as well as design compression techniques, depending on the function f, which will limit the maximum communication at a node. In this scheme, every node v will gather (sketches of) data from its children, then calculate a *sketch* of the total data, effectively compressing the data from the subtree under it. It will then pass this information to its parent node stored in the subtree rooted at v and forwards them to its predecessor. Note that, since v will be receiving sketches from its children that it does not have the resources to uncompress, this implies that the sketch must have a certain modularity property, such as the random Karp-Rabin fingerprints which use polynomials for dimensionality reduction. This is a challenge in a randomized setting if we do not assume shared randomness, which is hard to achieve. Starting with the work of Greenwald and Khanna [5], the relationship between sensor network data aggregation and sketching in data stream processing has been well established, and we expect our results involving sketching in either area to influence the other. We will also investigate the impact of factors such as the network topology, data value distribution, and node connectivity on the efficiency of our solutions.

Clustering the sensors is another avenue that we will consider. Due to the way G is formed, highly connected clusters in G can have strongly correlated behaviour in terms of their data values, and thus exhibit low entropy/high compressibility. Our goal is to develop and analyze efficient distributed clustering algorithms. The techniques that we will use for identifying and constructing these clusters include decentralized coordination algorithms, computational geometry techniques, randomized sketching techniques for data aggregation, and clustering algorithms relying on nearest-neighbors search and hashing techniques.

This work is closely related to the distributed leader election algorithms discussed above, as well as the computational geometry perspective for the clustering. Random walks and gossip-based algorithms can provide efficient ways of distributing data while avoiding interference between nodes and minimizing communication. The relationship between data dissemination and aggregation is yet to be investigated in detail. Our teams would like to study the relative difficulty of the two problems, and explore the relationship between their solutions.

2.2 Impact

The collaboration will result in a number of unifying themes in approaching network algorithms. Ultimately the theoretical work will improve the understanding and the use of complex networks.

Our team members have ongoing collaborations/contacts with several institutions which are leaders in the networking field such as AT & T Research, Yahoo Research, Cisco, HP Research and IBM Research. By arranging visits from researchers in these institutions, we expect to strengthen our ties with industry and research labs, validate our research problems, and provide exposure of our students to the big players in the field and to prospective employers.

Our research teams have collaborations/contacts with a large number of well-known researchers from Canada, US, Europe, and Israel. We expect to see closer ties with researchers from various parts of the world where theoretical networking is strong through visits and seminars.

One of our main goals is to reach out to students and to expose them to a research area which is fun and deep, with real impact on technologies that they are using every day. Our aim is to share the joy of algorithmic network research with our undergraduate and junior graduate students, as well as provide an environment for our more senior graduate students to interact with each other and with a varied set of established researchers in the area of algorithms.

3 The Teams

Our lead PIs are Petra Berenbrink, Funda Ergun, and Valerie King. Throughout this proposal, all names have been listed alphabetically.

Simon Fraser University (lead institution):

- Petra Berenbrink (associate professor). Petra Berenbrink finished her Ph.D at the University of Paderborn, Germany, in 2000. After that she spent 18 month as a postdoctoral research fellow at Warwick University in the UK. She joined SFU in the beginning of 2002. Her main research interests are randomized algorithms, distributed computing and the analysis of dynamic systems.

- Funda Ergun (associate professor). Dr. Ergun received her Ph.D. from Cornell U.; afterwards she was a postdoctoral researcher at the U. of Pennsylvania, a research scientist in the Networking Department of Bell Labs, Murray Hill, and a Schroeder Assistant Professor at Case Western Reserve U . before joining SFU. Her research interests lie in randomized algorithms/sampling, sensor networks, massive/streaming data, and reliable transmission.

- Cenk Sahinalp (professor and CRC Tier II). Dr Sahinalp received his Ph.D. in computer science in from U. of Maryland, College Park. He was a research scientist at Bell Labs, Murray Hill, an assistant professor with tenure at U. of Warwick, and an assistant professor at Case Western U. before joining SFU. His research interests include string algorithms, bioinformatics especially in the context of protein interaction networks, and network modelling.

University of Victoria (lead institution):

- Bruce Kapron (professor). Dr. Kapron received his Ph.D. from U. of Toronto. Afterwards, he was a visiting scientist at Carnegie Mellon University and a postdoctoral fellow at SFU. In 1993 he joined U. of Victoria. He has been a visiting researcher at Stanford. His research interests are in the design of secure protocols of networks, verification of security systems, anonymizing graph data from social networks, logic and verification.

- Valerie King (professor). Dr. King received her Ph.D. from University of California, Berkeley. Afterwards, she was a postdoctoral fellow at Princeton U. and a research scientist at NECI. In 1992 she joined the University of Victoria. She has been a visiting researcher at Hebrew University and Microsoft Research (SVC). Her research interests are randomized algorithms and data structures, distributed computation and lower bounds, with application to networks.

- Venkatesh Srinivasan (associate professor). Dr. Srinivasan received his Ph.D. in computer science from Tata Institute of Fundamental Research in Mumbai. Afterwards, he had postdoctoral positions at Institute for Advanced Study, DIMACS, and the Max Planck Institute for Informatics before joining UVic. His current interests are graph anonymization for social networks and random network generation models.

University of Calgary:

- Lisa Higham (professor). Dr. Higham received her Ph.D. from UBC in 1988. Her research interest are in distributed computing, randomized algorithms and multiprocessor algorithms, as well as fault tolerance.

- Philipp Woelfel (assistant professor). Dr. Woelfel received his Ph.D. from U. of Dortmund, Germany. Afterwards he was a postdoctoral fellow at the U. of Toronto before joining U. of Calgary. His areas of interest are the theory of distributed computing, randomized algorithms/data structures, and computational complexity.

University of British Columbia:

- Will Evans (associate professor). Dr. Evans received his Ph.D. from the U. of California, Berkeley, after which he was a postdoctoral researcher at UBC and an assistant professor at U. of Arizona. Afterwards he re-joined UBC. His research interests are in the areas of computational geometry, sensor networks, program compaction, information theory, and theoretical aspects of computation in general.

- Nick Harvey (assistant professor). Nick Harvey got his PhD from MIT, after which he was a postdoc at Microsoft and an assistant professor at the University of Waterloo. He joined UBC in Fall 2011. His areas of interest include combinatorial optimization, graph theory, and network algorithms.

- David Kirkpatrick (professor). Dr. Kirkpatrick got his Ph.D. from the U. of Toronto after which he taught at Cornell U. and SFU before joining UBC. He is a fellow of the Royal Society of Canada. His interests lie in computational complexity and computational geometry, especially as they relate to sensor networks. He has recently become interested in using probabilistic algorithms for parallel and distributed computation.

Advisory committee We propose an advisory committee with the following experts: James Aspnes (Professor, Yale U.), Pavol Hell (Professor, Simon Fraser U.), Anna Karlin (Professor, U. of Washington), David Lee (Director, Networking Research, HP Labs, previously vice president of Bell Laboratories and founding director of Bell Laboratories China).

4 Proposed Activities

- Three PIMS PDFs for two years, one PDF per lead institution.

- **Flagship event**: a workshop at SFU in Spring 2014 (tentatively April or May) with the theme "Massive Scale Modeling for Networks and Algorithms". This will be held at the Harbour Centre in downtown Vancouver. The event will contain many talks, tutorials, and a student forum. We expect highly respected senior researchers to attend, as well as students from North America and from other international locations. Here is a list of people that we are going to invite. Timothy Chan (U of Waterloo), Colin Cooper (King's College London), Graham Cormode (AT&T), Matt Devos (SFU), Michael Friedlander (UBC), Leonidas Guibas (Stanford), Piotr Indyk (MIT), Luis Goddyn (SFU), Ravi Kumar (Yahoo Research), Sampath Kannan (U. of Pennsylvania), Robert Krauthgamer (Weizmann Institute), Gary MacGillivray (U. of Victoria), Bojan Mohar (SFU), Dana Ron (Tel Aviv U.) Christian Sohler (U. of Dortmund), Eva Tardos (Cornell U), Ozgur Yilmaz (UBC).

Directly after the event we will have two summer mini-courses at SFU Burnaby over three days.

- A distinguished lecture series with twelve visitors from academia and research labs/industry in the participating venues over two years. Candidates are Ali Begen (Cisco), Artur Czumaj (University of Warwick), Mark Jerrum (U. of Edinburgh), Anna Karlin (University of Washington, confirmed), David Kempe (USC, confirmed), Ravi Kumar (Yahoo Research, confirmed), Ronitt Rubinfeld (MIT, confirmed), Rakesh Sinha (AT&T Research), Christian Sohler (Technical U. of Dortmund), Eli Upfal (Brown U.). Berthold Vöcking (RWTH Aachen U.), Lisa Zhang (Bell Laboratories).

  We expect six of these visitors to stay for a longer-term visit.

- A workshop at UVic on ”Randomization in networks”, where the PIs as well as students and researchers in the field will be able to share their work. Proposed date: October 2012. Again a list of researchers that we are going to invite. Bernard Chazelle (Princeton U.), Artur Czumaj, (Warwick U.), Erik Demaine (MIT), Alan Frieze (Carnegie Mellon U.), Friedhelm Meyer auf der Heide (U. of Paderborn), Christian Scheideler (U of Paderborn), Harald Räcke (TU Munich), Gabor Tardos (SFU), Berthold Voecking (Aachen U.), Lisa Zhang (Bell Laboratories).

  Two short courses (one to two weeks in the format of a summer school) in conjunction with the workshop described above.

- Longer term visits by the PIs and their highly qualified personnel between institutions.

- Collaboration with US institutions such as AT & T, Yahoo research, Cisco, HP Research, IBM, MIT, University of New Mexico (collaborators: Ali Begen (Cisco), Ravi Kumar (Yahoo), David Lee (HP), Kostas Oikonomou (AT&T), Kadangode Ramakrishnan (AT&T), Ronitt Rubinfeld (MIT), Jared Saia (UNM), Rakesh Sinha(AT& T), David Woodruff (IBM)).

5 Budget

- One PDF for each lead institution (three total PDFs) for two years each. $120,000. The participating institutions will cover the other half of their salaries as customary with this program and provide various other necessary support using sources such as NSERC Discovery Grants.

- Flagship Workshop, followed by two minicourses. Invited speakers, student travel grants, food/equipment. $10,000 for the workshop, $5000 for the minicourses. $15,000 overall.

- Twelve speakers, $1800 for four speakers coming from the East Coast, $1200 for four speakers from the West Coast, $2500 for four overseas speakers, covering travel costs. $22,000 overall..

  Six longer term visitors over two years, room and board costs, $18,000 overall. (SFU CS is estimated to match two of them with $3000 each.)

- One workshop at UVic, $8000.
- Travel between participating institutions and short and long term visits by the participants. $40,000.
- No explicit funds requested for US collaborations; matching funds from NSERC Discovery Grants of participants: $ 25,000/year, in addition to support for the PIMS PDFs and graduate student support by the PIs. The participating universities will contribute with visitor funds to the research visitors.

Grand total: $223,000 including PDF support.

References

[1] R. Albert; A.-L. Barabsi *Statistical mechanics of complex networks.* In Reviews of Modern Physics 74: 4797.

[2] Gurkan Bebek, Petra Berenbrink, Colin Cooper, Tom Friedetzky, Joseph H. Nadeau, Suleyman Cenk Sahinalp. *The degree distribution of the generalized duplication model.* In Theor. Comput. Sci. (TCS) 369(1-3):239-249 (2006).

[3] Colin Cooper, Alan M. Frieze. *The cover time of the preferential attachment graph.* In J. Comb. Theory, Ser. B 97(2): 269-290 (2007)

[4] Fereydoun Hormozdiari, Petra Berenbrink, Natasa Przulj, Suleyman Cenk Sahinalp. *Not All Scale Free Networks Are Born Equal.* The Role of the Seed Graph in PPI Network Emulation. Systems Biology and Computational Proteomics 2006: 1-13.

[5] Michael Greenwald, Sanjeev Khanna. *Power-Conserving Computation of Order-Statistics over Sensor Networks.* Principles of Database Systems (PODS), 2004.