Expansion Quality of Epidemic Protocols

Pasu Poonpakdee and Giuseppe Di Fatta

Abstract. Epidemic protocols are a bio-inspired communication and computation paradigm for large and extreme-scale networked systems. This work investigates the expansion property of the network overlay topologies induced by epidemic protocols. An expansion quality index for overlay topologies is proposed and adopted for the design of epidemic membership protocols. A novel protocol is proposed, which explicitly aims at improving the expansion quality of the overlay topologies. The proposed protocol is tested with a global aggregation task and compared to other membership protocols. The analysis by means of simulations indicates that the expansion quality directly relates to the speed of dissemination and convergence of epidemic protocols and can be effectively used to design better protocols.

Keywords: epidemic protocols, expander graphs, extreme-scale computing, decentralised algorithms.

1 Introduction

In extreme-scale distributed systems, computing and spreading global information is a particularly challenging task. Centralized paradigms are not suitable, as they introduce bottlenecks and failure intolerance; fully decentralised and fault-tolerant approached are very desirable.

Epidemic (or Gossip-based) protocols are a robust and scalable communication paradigm to disseminate information in a large-scale distributed environment using randomised communication [1]. The advantages of epidemic protocols over global

Pasu Poonpakdee · Giuseppe Di Fatta
School of Systems Engineering, University of Reading, Whiteknights,
Reading, Berkshire, RG6 6AY, United Kingdom
e-mail: p.poonpakdee@pgr.reading.ac.uk, g.difatta@reading.ac.uk

D. Camacho et al. (eds.), *Intelligent Distributed Computing VIII,*
Studies in Computational Intelligence 570, DOI: 10.1007/978-3-319-10422-5_31

communication schemes based on deterministic interconnection overlay networks are their inherent robustness and scalability.

Applications based on epidemic protocols are emerging in many fields, including Peer-to-Peer (P2P) overlay networks (e.g., [2]), mobile ad hoc networks (MANET) (e.g., [3]) and wireless sensor networks (WSN) (e.g, [4]). More recently, epidemic protocols have been adopted to support fully decentralised data mining tasks for extreme-scale systems [5, 6], even under node churn and network failures [7].

In general, Epidemic protocols can be adopted for information dissemination and to solve the distributed data aggregation problem in a fully decentralized manner. The goal of data aggregation in networks is the parallel determination at each node of the exact value, or of a good approximation, of a global aggregation function over a distributed set of values. In this work, global aggregation is used as a typical application for the performance analysis of epidemic protocols.

The expansion property of graphs is a fundamental mathematical concept [8], which has been adopted to study several aspects of complex networks and social graphs [9, 10], and quasirandom rumor spreading was shown to exhibit a natural expansion property [11].

This work investigates the expansion property of the overlay topologies induced by epidemic protocols. The expansion quality of epidemic membership protocols is introduced and simulations show that it is a good indicator of the convergence speed of an epidemic global aggregation.

The rest of the paper is organised as follows. Section 2 introduces the adopted definition and measures of the expansion property of graphs. In Section 3 epidemic protocols are briefly reviewed. Section 4 introduces a novel expander membership protocol. Section 5 presents the results of simulations and their analysis. Finally, Section 6 provides some conclusive remarks and directions of future work.

2 Expansion Property of Graphs

Expander graphs, or simply expanders, are sparse graphs that have strong connectivity properties. Informally, a graph is an expander if any vertex subset (not too large) has a relatively large set of one-hop distant neighbours.

There are a few alternative definitions of expanders, which are based on the vertex expansion, the edge expansion or the spectral gap. The definition of expanders adopted in this work, is based on the vertex expansion.

Given a graph $G = (V, E)$, where V is the set of nodes and E the set of edges, and a subset of nodes $S \subset V$, the outer boundary of S is the set of nodes that are not in S and have at least one neighbour in S. Specifically, the outer boundary is defined as:

$$\partial(S) = \{v \in V \backslash S : \exists u \in S \mid \langle u, v \rangle \in E\}. \quad (1)$$

The *expansion ratio* of a subset $S \subset V$ is defined as $\frac{|\partial(S)|}{|S|}$. Although the typical measure of expansion quality of a graph is based on the *expansion ratio*, it is sometimes convenient to adopt some normalised variant [9]. The relative size of the outer boundary of a subset $S \subset V$ is here adopted to define the **vertex expansion index** $h(G, S)$ as:

$$h(G, S) = \frac{|\partial(S)|}{|V \backslash S|}. \quad (2)$$

The vertex expansion index is defined in $[0, 1]$ regardless of the graph order ($|V|$) and the sample size ($|S|$), while the *expansion ratio* is not.

For a given sample size s, the minimum and maximum vertex expansion indices are defined as:

$$h_{min}(G, s) = \min_{S \subset V, |S|=s} \frac{|\partial(S)|}{|V \backslash S|} \; and \quad (3)$$

$$h_{max}(G, s) = \max_{S \subset V, |S|=s} \frac{|\partial(S)|}{|V \backslash S|}. \quad (4)$$

These two expansion indices measure the range of the outer boundary cardinality for a given sample size with respect to the largest possible outer boundary.

Expanders are typically characterised by the minimum value of the expansion property over a specific range of the sample size, i.e. $0 < |S| \leq \frac{|V|}{2}$ However, in this work, we have adopted a fixed sample size to carry out the analysis. In a preliminary analysis, we have experimentally determined that a sample size of 5% of the order of the graph is a good choice, as larger samples may reach the largest possible expansion.

3 Epidemic Protocols

Epidemic protocols are typically described as periodic and synchronous with a cycle length Δ_T and are executed for a number of cycles T_{max}. At all discrete times t ($0 < t \leq T_{max}$), each node independently sends information to a peer, which is ideally selected uniformly at random among all nodes in the system. This selection operation is provided by a peer sampling service, the *membership protocol*, which is implemented with an epidemic approach.

Epidemic membership protocols are necessary to support application-level protocols, which provide services such as decentralised data aggregation. At each cycle, a membership protocol defines an overlay topology, over which communication operations of the application-level service are based.

Membership protocols and aggregation protocols are briefly reviewed in the following sections.

3.1 Membership Protocols

The node sampling service is considered a fundamental abstraction in distributed systems [12]. In large-scale systems, nodes cannot build and maintain a complete directory of memberships. A membership protocol builds and maintains a partial view of the system, which is used to provide the random node selection service. The distributed set of views implicitly defines an overlay topology $G = (V, E)$. A membership protocol periodically and randomly changes the local views, thus generating a sequence of random overlay topologies $\Gamma = \{G_i\}$, with $G_i = (V_i, E_i)$ being the overlay topology at protocol cycle i.

The required assumptions are that the physical network topology is a connected graph, a routing protocol is available and an initialisation mechanism for the overlay topology is provided.

Several membership protocols have been proposed in the literature, which have typically been designed to produce random overlay topologies. However, none of the previous approaches has considered the expansion quality of the induced overlay topologies as design principle, nor to analyse the performance.

The *Node Cache Protocol* [13] is the simplest membership protocol, which adopts a straightforward approach based on a symmetric *push-pull* mechanism. At each node, the protocol maintains a local cache Q of node identifiers (IDs), with $|Q| = q_{MAX}$. At each protocol cycle, the content of the local cache is sent (*push*) to a node randomly chosen from the local cache according to a uniform probability. When a remote cache is received, the local cache is sent (*pull*) to the remote node. The remote cache and the remote node ID (refresh mechanism) are merged with the local cache, which is finally trimmed to the maximum size by randomly removing the number of IDs exceeding q_{MAX}. The node cache protocol provides a local service which approximates a random peer sampling in the global system. When invoked, the service removes and returns a random node from the local cache.

The protocol *Send&Forget* [14] is based on a simple *push* mechanism. At each protocol cycle, a portion of the local cache is sent to a node randomly chosen from the local cache according to a uniform probability. When a remote cache is received, it is merged with the local cache.

Cyclon [15] is a membership protocol that is an enhancement of a basic shuffling mechanism similar to the one adopted in the Node Cache Protocol. Cyclon adopts a lifetime (age) of the node IDs in the local cache and the selection of entries in the local cache is biased by their lifetime.

The protocol *Eddy* [16] attempts to minimize temporal and spatial dependencies between nodes' caches in order to provide a better random distribution of the node samples in the overall system. Eddy is arguably the most complex membership protocol and may incur in significant communication overhead.

3.2 Aggregation Protocols

The data aggregation problem refers to the computation of a global aggregation function in a network of nodes, where each node is holding a local value. Examples of global aggregation functions are the sum, the average, the maximum, the minimum, random samples, quantiles, etc. Local approximations of the global aggregate function can be obtained with an epidemic aggregation protocol.

Nodes periodically exchange their local state and the reception of a remote state triggers the update of the local state at a node. The update produces a reduction of the variance in the estimates in the system until convergence. The definitions of local state, type of messages and update operation depend on the particular aggregation protocol and the target global function. A good approximation of the global aggregate function can be obtained at every node within a number of protocol cycles.

Epidemic aggregation protocols may typically employ *push*, *pull* or *push-pull* schemes. The Symmetric Push-Sum Protocol (SPSP) [13] combines the accuracy and simplicity of a push-based approach and the efficiency of the push-pull scheme. SPSP does not require synchronous communication with atomic operations; it achieves a convergence speed similar to the push-pull scheme, while keeping the accuracy of the push scheme. SPSP and the global average as target function have been used in the simulations presented in this work.

4 An Expander Membership Protocol

The membership protocol introduced in this section is the first attempt to directly exploit the concept of expansion in graphs.

Memberships protocols can be seen as a distributed implementation of multiple random walks and are used to generate random overlay topologies that are sparse and have strong connectivity. After sufficiently many protocol cycles the set of neighbours (or outgoing edges) of each node are expected to be uniformly distributed. A quick convergence to a random overlay topology is an important quality of membership protocols, as it affects the convergence speed of the applications.

These considerations have inspired a new membership protocol based on the concept of vertex expansion, as briefly outlined here. The protocol is based on the symmetric (*push-pull*) cache shuffling approach as described in section 3.1. At each cycle and at each node i, a destination node x_0 is randomly selected from the local cache Q_i. A *push* message m_i containing the local cache Q_i is sent to x_0. At the reception of the message, node x_0 computes the intersection of the local cache Q_{x_0} and the remote cache Q_i. The message is accepted if $|Q_{x_0} \cap Q_i| \leq T_{max}$, where T_{max} is a neighbourhood similarity threshold ($T_{max} \geq 0$). If the incoming *push* message is accepted, a *pull* message is sent to node i and the two caches are merged. Otherwise, $|Q_{x_0} \cap Q_i| > T_{max}$ and the message m_i is forwarded to a node x_1 randomly selected from the local cache of node x_0. This procedure is repeated up to a maximum number of hops (H_{max}).

This simple protocol aims at maximising the expansion quality of the overlay topology by swapping and merging cache entries between nodes with low neighbourhood similarity. Therefore, clusters of nodes with high neighbourhood similarity are expected to break up sooner than in the 1-hop push-pull scheme.

This protocol may require additional components to optimise other aspects of the membership management task. However, for the purpose of this work this membership protocol is suitable to show the relation between expansion quality and the convergence speed of a global aggregation task.

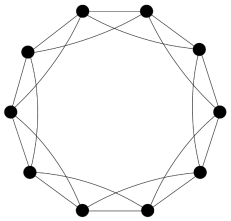


Fig. 1 Initial overlay topology: a regular circular lattice

5 Experimental Analysis

The goal of the experimental analysis is to compute measures of the expansion quality of the overlay topologies induced by different membership protocols and to investigate their relation with the convergence speed of an epidemic aggregation task in the network.

A membership protocol implements a distributed and continuous edge rewiring procedure to generate random overlay graphs. Edge rewiring is performed by means of random communication operations over the current overlay topology.

When a membership protocol is executed over a random regular graph, it generates a sequence of random regular graphs with similar characteristics. In this case, all membership protocols seem to provide an acceptable peer sampling service with respect to the convergence speed of the global aggregation. However, when the overlay topology is not a regular random graph, differences between protocols emerge. This may happen, for example, when the overlay topology is initialised (cold start) or when high churn introduces perturbations. Rather than studying optimal initialisation procedures, here we investigate the ability of a protocol to change a poor topology into a random graph quickly. This property is intuitively connected with the expansion quality of graphs, hence the concept of the expansion quality of a membership protocol.

A poor topology could be generated by a naive centralised initialisation procedure, e.g. a star topology, a small clique of servers to which all nodes are connected, or a circular lattice. They all have a poor expansion quality, and without the random rewiring mechanism of a membership protocol they would lead to a poor

performance of global aggregation and information dissemination protocols. Obviously, centralised topologies also suffer from load imbalance. In the experimental analysis, the overlay topology is initialised as circular lattice (Figure 1) with a constant out degree (30), which provides a poor expansion with a good load balance.

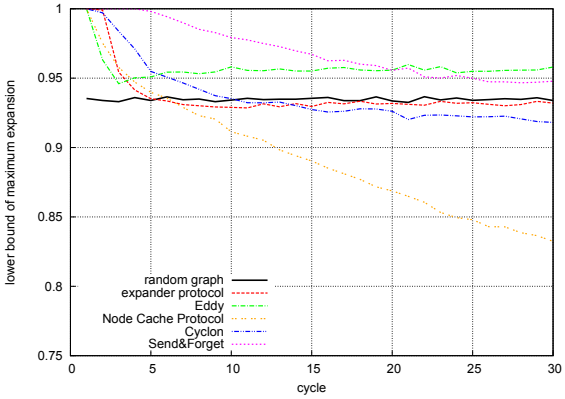
An asynchronous network configuration with a uniform distribution of network latencies has been adopted. The simulations have been executed with the following membership protocols, whose parameters have been set for best performance according to the literature and to a preliminary analysis:

- *Node Cache Protocol* [13] ($q_{max} = 30$),
- *Send&Forget* [14] with cache size upper and lower bounds of, respectively, 40 and 15,
- *Cyclon* [15] with shuffle length of 15,
- *Eddy* [16] with refresh rate of 10 cycles and shuffle length of 15,
- the novel expander protocol ($q_{max} = 30$, $H_{max} = 5$ and $T_{max} = 0$) and
- an ideal protocol based on random graphs with a constant outdegree of 30.

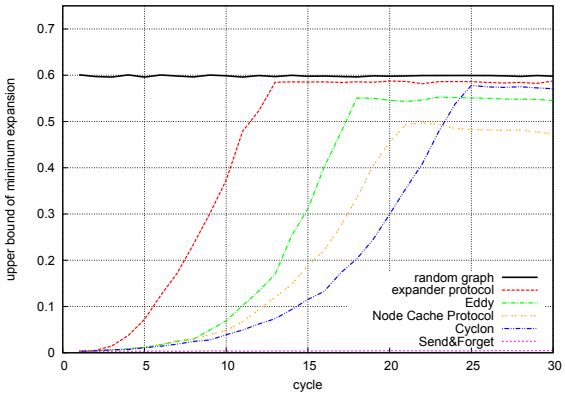
In all protocols the initial cache size (outdegree) is set to 30. The random membership protocol is included to provide a baseline performance and is based on the ideal global knowledge of the system to generate a different random graph at each cycle of the protocol.

In the first set of simulations (Figure 2), each protocol has been run for a number of cycles starting from the initial circular lattice topology. The expansion indices h_{min} and h_{max} for a sample size of 5% are used to monitor the evolution of the expansion quality of the overlay topology over the cycles. The exact values of the indices cannot be determined, as they would require an exhaustive search over a combinatorial number of node subsets. The approximation of the minimum and maximum expansion indices can be determined by a greedy algorithm. The adopted greedy algorithm is a variant of the one adopted in [9], where at step 6 the objective function $|N(S \cup \{v\})|$ is used in place of $|N(\{v\}) - (N(S) \cup S)|$. The different function can provide a tighter bound of the extreme values (min/max) when $v \in N(S)$. Although the values determined by the greedy method are just upper and lower bounds of the true extreme values, they are believed to be a good approximation.

In the charts of Figure 2, the minimum and maximum values of the vertical axis have been chosen to correspond to the average expansion index ($\bar{h} \approx 0.77$), which was determined by a Montecarlo method. Figure 2 shows that the maximum expansion index of the expander protocol converges quickly to the one of the random protocol. *Eddy* has a higher maximum and a lower minimum than the random protocol. The minimum expansion index of the expander protocol shows the fastest rate of convergence to the one of the random protocol, followed by *Eddy*, *Node Cache Protocol* and *Cyclon*. The minimum expansion index of *Send&Forget* remains close to 0 for the entire range of protocol cycles: in this case the overlay topology is not changing fast enough from the initial ring lattice. This test shows that the protocol that is explicitly designed to optimise the expansion quality of the overlay topology, as expected, achieves this goal better than the other membership protocols.



(a) maximum expansion index h_{max}



(b) minimum expansion index h_{min}

Fig. 2 Greedy approximation of the expansion indices (network size: 10000 nodes, sample size: 5%)

In the second set of simulations (Figure 3), the aggregation protocol *SPSP* [13] has been adopted to perform a global aggregation. The local values of the aggregation protocol are initialised with a peak distribution: all nodes have initial value 0, but one that has a peak value. After some protocol cycles the local values are expected to converge to the expected target value (global average). The standard deviation of the local aggregation values is used to measure the convergence speed of the aggregation protocol with the different membership protocols.

Figure 3 shows how the membership protocol can be relevant in terms of the convergence speed of a global aggregation task. The random protocol has a constant slope, which corresponds to a constant rate of the variance reduction in the system. The other membership protocols need to change the initial topology into a random graph before they can also provide a similar variance reduction rate (when the curves become parallel to the one of the random protocol). It is evident that the protocol explicitly based on the concept of expansion has the best performance. The protocol

Send&Forget has a particularly poor performance: it takes a very long time to rewire the topology into a random graph because it has an asymmetric communication pattern (*push* only) and rewires only a small number of edges for each message (shuffle length is 2).

The maximum expansion index (Figure 2(a)) does not seem to be a good indicator of the quality of the membership protocol. While the minimum expansion index (Figure 2(b)) is rather interesting: it clearly provides the same ranking of protocols as in Figure 3.

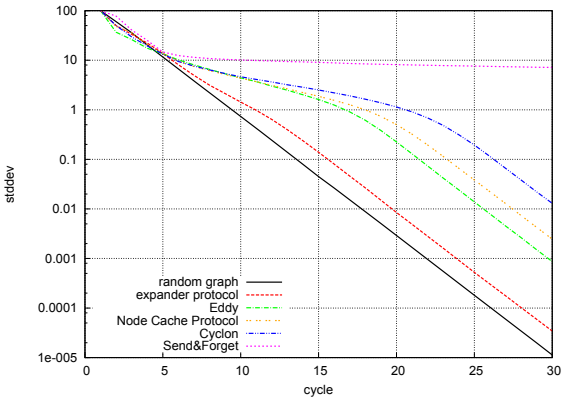


Fig. 3 Convergence speed of epidemic aggregation (network size: 10000 nodes)

Finally, the effect of different values for the parameters H_{max} and T_{max} of the expander protocol has been analysed, although the detailed results are not reported for the sake of brevity. T_{max} has a significant effect in the performance, with values closer to zero being the most effective. The effect of H_{max} is also important, though for $H_{max} > 2$ the improvement is less evident.

This work has shown that the minimum expansion index is useful to characterise epidemic membership protocols. Although evidence from more extensive simulations and experimental analysis in real-world networks would be required for more conclusive results.

6 Conclusions

This work has investigated how the convergence speed of decentralised epidemic aggregation may be affected by the underlying membership protocol. Membership protocols have the task of generating and evolving random overlay topologies to support fast epidemic aggregation and information dissemination. When the overlay topology is far from random, the membership protocol is expected to rewire the edges in such a way to quickly converge to random graphs. It has been shown that

different membership protocols perform this task at quite different rates, and, most importantly that the minimum expansion quality is a good candidate as indicator of this difference. Future work will be devoted to extend the experimental analysis to other initial topologies and network conditions, and in particular to study the effect of node churn.

References

1. Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., Terry, D.: Epidemic algorithms for replicated database maintenance. In: Proc. of the Sixth Annual ACM Symposium on Principles of Distributed Computing, PODC 1987, pp. 1–12. ACM (1987)
2. Ghit, B., Pop, F., Cristea, V.: Epidemic-style global load monitoring in large-scale overlay networks. In: International Conference on P2P, Parallel, Grid, Cloud, and Internet Computing, pp. 393–398 (2010)
3. Ma, Y., Jamalipour, A.: An epidemic P2P content search mechanism for intermittently connected mobile ad hoc networks. In: IEEE GLOBECOM, pp. 1–6 (2009)
4. Chitnis, L., Dobra, A., Ranka, S.: Aggregation methods for large-scale sensor networks. ACM Transactions on Sensor Networks (TOSN) 4, 1–36 (2008)
5. Di Fatta, G., Blasa, F., Cafiero, S., Fortino, G.: Epidemic k-means clustering. In: Proc. of the IEEE Int.l Conf. on Data Mining Workshops, pp. 151–158 (2011)
6. Mashayekhi, H., Habibi, J., Voulgaris, S., van Steen, M.: GoSCAN: Decentralized scalable data clustering. Computing 95(9), 759–784 (2013)
7. Di Fatta, G., Blasa, F., Cafiero, S., Fortino, G.: Fault tolerant decentralised k-means clustering for asynchronous large-scale networks. Journal of Parallel and Distributed Computing 73(3), 317–329 (2013)
8. Hoory, S., Linial, N., Wigderson, A.: Expander graphs and their applications. Bullettin of the American Mathematical Society 43(4), 439–561 (2006)
9. Maiya, A.S., Berger-Wolf, T.Y.: Expansion and search in networks. In: Proc. 19th ACM Intl. Conference on Information and Knowledge Management, CIKM 2010 (October 2010)
10. Malliaros, F.D., Megalooikonomou, V.: Expansion properties of large social graphs. In: DASFAA International Workshop on Social Networks and Social Media Mining on the Web (SNSMW) (April 2011)
11. Doerr, B., Friedrich, T., Sauerwald, T.: Quasirandom rumor spreading: Expanders, push vs. pull, and robustness. In: Albers, S., Marchetti-Spaccamela, A., Matias, Y., Nikoletseas, S., Thomas, W. (eds.) ICALP 2009, Part I. LNCS, vol. 5555, pp. 366–377. Springer, Heidelberg (2009)
12. Jelasity, M., Voulgaris, S., Guerraoui, R., Kermarrec, A.M., van Steen, M.: Gossip-based peer sampling. ACM Trans. Comput. Syst. 25(3) (2007)
13. Blasa, F., Cafiero, S., Fortino, G., Di Fatta, G.: Symmetric push-sum protocol for decentralised aggregation. In: Proc. of the Int.l Conf. on Advances in P2P Systems, pp. 27–32 (2011)
14. Gurevich, M., Keidar, I.: Correctness of gossip-based membership under message loss. In: Proceedings of the 28th ACM Symposium on Principles of Distributed Computing, PODC 2009, pp. 151–160. ACM, New York (2009)
15. Voulgaris, S., Gavidia, D., Steen, M.: Cyclon: Inexpensive membership management for unstructured p2p overlays. Journal of Network and Systems Management 13(2), 197–217 (2005)
16. Ogston, E., Jarvis, S.A.: Peer-to-peer aggregation techniques dissected. Int. J. Parallel Emerg. Distrib. Syst. 25(1), 51–71 (2010)