

Bounding the convergence time of local probabilistic evolution

Simon Apers, Francesco Ticozzi, Alain Sarlette

► **To cite this version:**

Simon Apers, Francesco Ticozzi, Alain Sarlette. Bounding the convergence time of local probabilistic evolution. Frank Nielsen; Frédéric Barbaresco. Geometric Science of Information, Second International Conference, GSI 2017, Nov 2017, Paris, France. Springer, 10589, LNCS - Lecture Notes in Computer Science. <hal-01634611>

HAL Id: hal-01634611

<https://hal.inria.fr/hal-01634611>

Submitted on 14 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bounding the convergence time of local probabilistic evolution

Simon Apers¹, Alain Sarlette^{1,2}, and Francesco Ticozzi^{3,4}

¹ Department of Electronics and Information Systems, Ghent University, Belgium

² QUANTIC lab, INRIA Paris, France

³ Dipartimento di Ingegneria dell'Informazione, Università di Padova, Italy

⁴ Department of Physics and Astronomy, Dartmouth College, NH 03755, USA

simon.apers@ugent.be, alain.sarlette@inria.fr, ticozzi@dei.unipd.it

Abstract. Isoperimetric inequalities form a very intuitive yet powerful characterization of the connectedness of a state space, that has proven successful in obtaining convergence bounds. Since the seventies they form an essential tool in differential geometry [1, 2], graph theory [4, 3] and Markov chain analysis [8, 5, 6]. In this paper we use isoperimetric inequalities to construct a bound on the convergence time of any local probabilistic evolution that leaves its limit distribution invariant. We illustrate how this general result leads to new bounds on convergence times beyond the explicit Markovian setting, among others on quantum dynamics.

This paper is concerned with the discrete-time spreading of a distribution along the edges of a graph. In essence we establish that even by exploiting global information about the graph and allowing a very general use of this information, this spreading can still not be accelerated beyond the conductance bound. Before providing more ample context, we start with a motivating example ascribed to Eugenio Calabi, but which came to our attention through the seminal 1969 paper by Jeff Cheeger [1]. Whereas the original example concerns differential geometry, we will apply it to a graph setting.

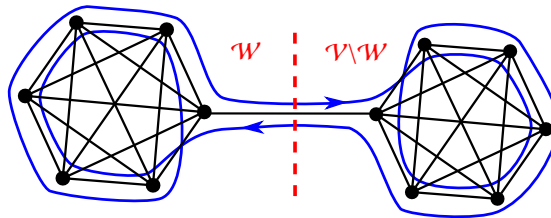


Fig. 1. (in black) Dumbbell graph K_n-K_n for $n = 6$. (in blue) superimposed cycle of length $4n$ in a construction towards faster mixing.

Consider a locality structure (discrete geometry) prescribed by the “dumbbell” graph family $K_n - K_n$ shown in Figure 1, consisting of two complete graphs over n nodes, connected by a single edge. The diameter of this graph is three, but a random walk over

this graph converging to the uniform distribution has an expected convergence time in $O(n^2)$. This convergence time can be improved with a “global design” but without violating locality of the evolution, by adding some memory to the walker. In Figure 1 (in blue), the system designer has superimposed a cycle over the dumbbell graph. By adding subnodes that allow to conditionally select different subflows through the graph (formally we “lift” the walk [9]), the walker can be restricted to walk along this cycle. Using this cycle, we can impose a strategy by Diaconis, Holmes and Neal [18, 9] to efficiently speed-up mixing over this cycle: *let the walker cycle in the same direction with a probability $1 - 1/n$, and switch direction with probability $1/n$* . This way the walk will mix over the graph in $O(n)$, i.e. quadratically faster than the original random walk. But this is still order n times slower than the diameter. Nevertheless, we show in our paper that this improvement is the best possible for any local probabilistic process that leaves the target distribution invariant. So mixing in diameter time may be possible, but not without loosening any of these constraints.

1. Problem description and main result: Consider a graph \mathcal{G} with nodes \mathcal{V} and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We use the convention that $(i, i) \in \mathcal{E} \forall i \in \mathcal{V}$. We define “states” X as probability distributions over \mathcal{V} . Given an initial state X_0 , some system “ \rightarrow ” propagates it over t time steps as $X_0 \xrightarrow{t} X_t$. For a subset $\mathcal{W} \subseteq \mathcal{V}$ and a state X , we define $X(\mathcal{W})$ the probability of \mathcal{W} according to X , and $X|\mathcal{W}$ as the state X conditioned on being in \mathcal{W} . We call $\mathcal{N}(\mathcal{W})$ the neighborhood of $\mathcal{W} \subseteq \mathcal{V}$, i.e., the nodes outside \mathcal{W} that have an edge going to \mathcal{W} . We impose the following fundamental properties.

- **linear initialization:** $X_0 \xrightarrow{t} X_t, \tilde{X}_0 \xrightarrow{t} \tilde{X}_t \Rightarrow pX_0 + (1-p)\tilde{X}_0 \xrightarrow{t} pX_t + (1-p)\tilde{X}_t$
- **locality:** $\forall X_0, t \geq 0, \mathcal{W} \subseteq \mathcal{V} : X_{t+1}(\mathcal{W}) \leq X_t(\mathcal{W}) + X_t(\mathcal{N}(\mathcal{W}))$
- **invariance:** $X_0 \xrightarrow{+\infty} \Pi \forall X_0 \Rightarrow \Pi \xrightarrow{t} \Pi$

The last property states that the unique steady state distribution of the system must be invariant as an initial condition. The second property expresses that probability weight can only flow along an edge at each time, without referring to details of the system mapping “ \rightarrow ”. The first property is natural as the input is a probability distribution. The point however is that the general process “ \rightarrow ” may e.g. contain hidden states, and we here impose a linear initialization with the hidden states as well (see example below).

Our theorem presents a bound on the convergence of a system “ \rightarrow ” that obeys these conditions towards its steady state Π . Explicitly, let τ be a time step such that $\|X_t - \Pi\|_1 \leq 1/2$ for all $t \geq \tau$. In discrete geometry, given a graph \mathcal{G} and a limit distribution Π , the isoperimetric measure Φ , which we also call the “conductance” [20], can be defined as:

$$\Phi = \max_P \Phi(P), \quad \Phi(P) = \min_{\mathcal{W} \subseteq \mathcal{V} : \Pi(\mathcal{W}) \leq 1/2} [P \circ (\Pi|\mathcal{W})](\mathcal{V} \setminus \mathcal{W}).$$

The maximization is over all stochastic matrices P acting on $\mathbb{R}^{|\mathcal{V}|}$ that obey the locality of \mathcal{G} and for which $P \circ \Pi = \Pi$. In other words, “ $P \circ$ ” is the most basic type of system “ \rightarrow ” satisfying our requirements: it is time-invariant and memoryless (“Markov”). If Π is the uniform distribution, then Φ is upper bounded by the edge expansion of \mathcal{G} , which is $1/n$ for the dumbbell graph. We establish the following “conductance bound” for any more complicated system.

Theorem 1. *If a system is linear, local and invariant, then $\tau \geq 1/(8\Phi)$.*

So for the dumbbell graph with Π the uniform distribution, we find $\tau \geq n/8$ for any linear, local and invariant system.

Mixing on graph structures has drawn much interest for sampling algorithms, see e.g. perfect matching [16], or the Metropolis-Hastings algorithm used a lot in statistical mechanics. Bounds similar to Thm.1 have originally been proven by Cheeger [1] and Buser [2] in a differential geometry setting, and by Fiedler [4], Dodziuk [8] and Alon [3] in a discrete geometry and graph setting. In Markov chain analysis, early uses trace back to Aldous [5], Lawler and Sokal [6] and Mihail [7]. More recent examples are by Chen, Lovász and Pak [9] who used a similar bound to prove that a restricted class of extended Markov chains called “lifted Markov chains” can at most quadratically accelerate convergence, and by Aharonov et al. [10] to (loosely) bound the convergence speed of certain quantum processes.

Our result allows to improve known mixing bounds, e.g. for quantum processes, and to generalize bounds beyond usual Markov chain settings, e.g. by including nonlinear decision rules. Examples are briefly discussed after the proof.

2. Proof: Our proof essentially comes down to two steps:

- *locality implies particular simulability:* the locality condition implies that the dynamics can always be described using (time- and state-dependent) local stochastic matrices. This is not entirely trivial in such generality.
- *bound for extended Markov chains:* we rather straightforwardly combine these matrices in an extended Markov chain model, for which we can prove the bound along standard lines.

2.1. Locality implies simulability: A stochastic matrix P is local if the system $X_0 \rightarrow X_t = P^t \circ X_0$ is local. It is not hard to check that this coincides with the traditional definition, where locality means $P_{i,j} = 0$ whenever $(i, j) \notin \mathcal{E}$. The following Lemma kind of proves the converse. Its proof is inspired by a related result of Scott Aaronson [19], establishing the lemma for quantum systems whose evolution is governed by a local unitary matrix.

Lemma 1. *If “ \rightarrow ” is a local system, then for every pair (X_0, t) with $t > 0$ there exists a local stochastic matrix $P(t, X_0)$ such that $X_0 \xrightarrow{t} X_t = P(t, X_0) \circ X_{t-1}$.*

Proof. Call $Y = X_{t-1}$ and $Z = X_t$. We make a digression to *flows over capacitated networks* [11] and consider the one shown in Figure 2. The network consists of a source node s , a sink node t , and two copies of the graph nodes \mathcal{V} and \mathcal{V}' . Node s is connected to any node $i \in \mathcal{V}$ with capacity $Y(i)$, any node $i \in \mathcal{V}$ is connected with capacity 1 to any node $j \in \mathcal{V}'$ iff $(i, j) \in E$ (else the nodes are not connected), and any node $j \in \mathcal{V}'$ is connected to node t with capacity $Z(j)$. If this network can route a steady flow of value 1 from node s to node t , then the fraction of $Y(i)$ that is routed from $i \in \mathcal{V}$ towards $j \in \mathcal{V}'$ directly defines $P_{j,i}(t, X_0)$.

The max-flow-min-cut theorem [11] states that the maximum steady flow which can be routed from node s to node t is equal to the minimum cut value of the graph, where

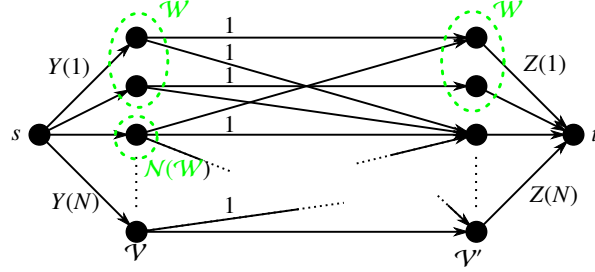


Fig. 2. Capacitated network construction used in Lemma 1.

a cut value is the sum of the capacities of a set of edges that disconnects s from t . It is clear that cutting all edges arriving at t disconnects the graph, with a cut value of 1, whereas cutting any middle edge between \mathcal{V} and \mathcal{V}' gives a cut value ≥ 1 . So the minimum cut need involve no such middle edge. Let us try to not cut the edges from $\mathcal{W} \subseteq \mathcal{V}$ to t . To block any flow from s to t while keeping all middle edges, we must then cut the edges from s to all the $l \in \mathcal{V}$ which have an edge to \mathcal{W} . This corresponds to all $l \in \mathcal{W} \cup \mathcal{N}(\mathcal{W})$. The value of this cut is thus

$$1 - \sum_{j \in \mathcal{W}} Z(j) + \sum_{j \in \mathcal{W}} Y(j) + \sum_{j \in \mathcal{N}(\mathcal{W})} Y(j) .$$

Recalling that $Y = X_{t-1}$ and $Z = X_t$, locality imposes

$$\sum_{j \in \mathcal{W}} Z(j) \leq \sum_{j \in \mathcal{W}} Y(j) + \sum_{j \in \mathcal{N}(\mathcal{W})} Y(j) ,$$

from which follows that the minimum cut value is ≥ 1 . According to the previous arguments, this concludes the proof. \square

2.2. Bound for “extended” Markov chains: On the basis of these $P(t, X_0)$, we show how to construct a local Markov chain with at most twice the convergence time τ of our original system “ \rightarrow ”. Thereto, we first define a closely related system

$$\rightsquigarrow \equiv \text{iterate } \xrightarrow{\tau} \text{ (namely } \text{floor}(t/\tau) \text{ times, plus } t \bmod \tau \text{ steps of it) .}$$

By construction, “ \rightsquigarrow ” has the same convergence time τ as “ \rightarrow ”, it has the same limit and it obeys the same locality and invariance conditions. We will now build a standard, time-invariant Markov chain that simulates the system “ \rightsquigarrow ”.

To this end we first extend our state space: the original node set \mathcal{V} is lifted to $\hat{\mathcal{V}} = (\mathcal{V}; \{0, \dots, \tau - 1\}; \mathcal{V})$. From the perspective of a random walker, the first item contains its starting position, the second item a clock variable, and the last item its current position. In matrix form, with \otimes representing the Kronecker product and \cdot^\dagger the transpose, we now build the transition matrix M for a Markov chain on $\hat{\mathcal{V}}$ as follows:

$$M = \sum_{i \in \mathcal{V}} \sum_{t=0}^{\tau-2} e_i e_i^\dagger \otimes e_{t+1} e_t^\dagger \otimes P(t, e_i) + \sum_{i, j \in \mathcal{V}} e_j e_i^\dagger \otimes e_0 e_{\tau-1}^\dagger \otimes e_j e_j^\dagger P(\tau - 1, e_i) .$$

Here e_i is the unit vector with 1 at index i , and $P(t, e_i)$ denotes the transition matrix obtained by Lemma 1 for $X_0 = e_i$, i.e., initial weight concentrated at node $i \in \mathcal{V}$. This Markov chain simulates the “ \rightsquigarrow ” system in the following sense: when we locally initialize it in the state

$$v[X_0] = \sum_{i \in \mathcal{V}} X_0(i) e_i \otimes e_0 \otimes e_i ,$$

the distribution over the subsets $(\mathcal{V}; \{0, \dots, \tau - 1\}; i)$ of the resulting state $M^t v[X_0]$ at time t exactly corresponds to X_t resulting from $X_0 \rightsquigarrow X_t$. A priori our Markov chain only simulates “ \rightsquigarrow ” for special initial states of the form $v[X_0]$ over $\hat{\mathcal{V}}$. The following lemma shows that in fact when starting from an arbitrary distribution over $\hat{\mathcal{V}}$, it takes at most twice the time to converge to $\|X_t - \Pi\|_1 \leq 1/2$ (over sets as just mentioned).

Lemma 2. *If \rightarrow has a convergence time τ , then the Markov chain M on $\hat{\mathcal{V}}$ has a convergence time at most 2τ over the subsets $\{(\mathcal{V}; \{0, \dots, \tau - 1\}; i) : i \in \mathcal{V}\}$.*

Proof. Consider an arbitrary initial state $e_i \otimes e_T \otimes e_k$ for the Markov chain M . After $\tau - T$ steps, this state will necessarily have evolved to one of the special initial states of the form $v[X_0]$, for some X_0 . By construction, the distribution of this state over the subsets $\{(\mathcal{V}; \{0, \dots, \tau - 1\}; i) : i \in \mathcal{V}\}$ will then simulate the evolution $X_0 \rightsquigarrow X_t$, which converges to $\|X_t - \Pi\|_1 \leq 1/2$ for all $t \geq \tau$. Note indeed that \rightsquigarrow and \rightarrow are equivalent over the first τ time steps, and furthermore by invariance, iterating \rightarrow via \rightsquigarrow will never increase $\|X_t - \Pi\|_1$. Hence the Markov chain will have converged after $\tau - T + \tau \leq 2\tau$ steps at most. We have thus proved the convergence time for initial states with all weight concentrated on one element of $\hat{\mathcal{V}}$. By linearity, this also proves the convergence time for arbitrary initial states. \square

The last element of the proof is a lower bound on τ for standard Markov chains. It essentially follows from a result by Chen, Lovász and Pak [9], stating that a class of extended Markov chains called “lifted Markov chains” that converge over $\hat{\mathcal{V}}$ can converge at best in order $1/\Phi$, with Φ the conductance of the original graph. Our Markov chain M does not exactly fit into this framework, because it is periodic on $\hat{\mathcal{V}}$ and only its projection onto \mathcal{V} via the subsets of Lemma (2) will converge. The proof can however be adapted to this case. Due to space constraints we must refer the reader to [22] for a detailed proof, and we here only provide the statement:

Lemma 3. *The convergence time of M over the sets $\{(\mathcal{V}; \{0, \dots, \tau - 1\}; i) : i \in \mathcal{V}\}$ is lower bounded by $1/(4\Phi)$.*

By combining Lemma 2 and Lemma 3, we obtain that 2τ must be larger than $1/(4\Phi)$ and hence τ larger than $1/(8\Phi)$, as stated in the main theorem.

3. Examples: We now discuss a few examples to illustrate the generality of our result. Note that the mathematical result is not restricted to cases where X_t represents a probability distribution. It can apply to any situation where X_t remains positive, bounded and preserves the sum of its components. Such dynamics can appear in flow dynamics and e.g. average consensus algorithms for weight distribution [13]. In such settings our result might suggest how e.g. relaxing the linearity constraint is necessary for beating the conductance bound.

3.1 Time-inhomogeneous Markov chains and Cesaro mixing: The bound clearly includes time-varying Markov chains (that satisfy invariance), as appear in the proof. Practical examples of such processes can be found in [23], and in [24] for card shuffling. The difficulty to analyze the convergence time of such processes is explicitly stated. Our paper thus provides a clear bound on the maximal achievable acceleration by exploiting the time-inhomogeneity degree of freedom in mixing algorithms.

Cesaro mixing [21] using a stochastic matrix P is defined by the system $X_0 \xrightarrow{t} X_t = \frac{1}{t+1} \sum_{k=0}^t P^k X_0$. There appears to be no obvious way to write this as a Markov chain. However, one can show that Cesaro mixing satisfies our assumptions, so Thm.1 allows to directly bound the mixing time of such processes.

3.2 Processes with state-dependent or nonlocal decision rules: The locality condition concerns how much probability weight is transferred at each step, but not how the decision about this transfer is taken. The latter is constrained by linearity. In this sense, our result directly bounds any attempts at adapting fast converging nonlinear algorithms e.g. from consensus, towards truly probabilistic Markov chains where linearity is natural. Consider a nonlinear update rule from consensus, like:

$$X_{t+1} = P(Z_t) X_t.$$

In [14], Z_t is a static function of the weight differences on the respective links, e.g. the weight associated to link (i, j) in $P(Z)$ is a function of $X_t(i) - X_t(j)$. Our framework would even admit Z_t being a dynamic function of X , possibly nonlinear, taking values in any space, and *would not even require that it is based on local values of X only*: e.g. the weight associated to link (i, j) in $P(Z)$ might be a function of some $X_t(k)$ where node k is totally elsewhere in the graph.

Such update rule is in general not linear in X_0 , but one may attempt to adapt it in this sense in the hope of designing e.g. stochastic automata that improve mixing over standard Markov chains. For instance, one might imagine a system that distinguishes, in memory, each part of X_t that has started from a different node at X_0 . Once this is done, we are free to choose the evolution (possibly nonlinear, nonlocal) for each of these X_0 -indexed parts, postulating that the full X_t consists of their linear combination. One might thus wonder whether such heuristic approach could lead to faster mixing on e.g. the dumbbell graph. Our result implies that — provided also invariance is required — such acceleration attempts are all limited to the conductance bound.

3.3 Finite-time convergence: Consider the following algebraic problem, related to finite-time convergence [25] and the inverse eigenvalue problem [26]:

What is the minimal number of doubly-stochastic matrices over a graph \mathcal{G} whose product has all but one eigenvalue different from zero?

A product of doubly-stochastic matrices over a graph \mathcal{G} describes a linear, local and invariant system, converging to the uniform distribution. If their product has all but one eigenvalue different from zero, then necessarily the system has converged. By our theorem the number of matrices is thus bounded by $1/(8\Phi)$.

3.4 Quantum walks: The convergence properties of quantum processes spreading over localized state spaces play a role both in physics (e.g. transport of excitations in photosynthesis [12]) and in quantum computation (e.g. quantum random walks [10]).

A discrete-time quantum walk is (although to our knowledge this has never been said explicitly) the generalization of a lifted walk, by keeping coherences among the node options. Denote by ρ the quantum state, i.e. a positive definite “density matrix” with trace one, whose diagonal represents probabilities over $\hat{\mathcal{V}} = \{(i, z)\}$ where $i \in \mathcal{V}$ is a graph node and z is a possible auxiliary degree of freedom (see introductory example [18], coined quantum walks [10], or Section 2.1). A general quantum walk follows:

$$\rho_{t+1} = \Phi \circ \rho_t$$

where Φ is a completely positive trace-preserving map; most popular is the unitary quantum walk, where $\Phi \circ \rho_t = U\rho_t U^\dagger$, with U a unitary matrix satisfying the locality of the graph \mathcal{G} , exactly as P does for a Markov chain and M does in Section 2.1. If ρ_t would remain diagonal, this would correspond exactly to a lifted Markov chain [9]. Authors have been wondering for some time whether the additional information contained in the off-diagonal elements of ρ_t (“quantum coherences”) might allow faster mixing. In [10] a conductance bound is given for *unitary* quantum walks and within a factor of the graph degree; but such factor becomes dominant for e.g. the dumbbell graph.

As will be further worked out in a future publication, quantum walks do satisfy the conditions of this paper, for most reasonable initializations. Then our result improves the bound of [10], both by generalizing it to non-unitary walks and by getting rid of the degree-dependent factor. Quantum walks indeed satisfy locality, including the hidden (complex) variables representing coherences. Linearity trivially holds, except if one allows to initialize the walk with nonlocal coherences already. In other words, since ρ_0 would necessarily be block-diagonal when all the initial weight is concentrated on a single node, introducing off-diagonal initial blocks when starting with a distribution over nodes would break linearity — and by Thm.1, this would be necessary to potentially beat the conductance bound.

References

1. Cheeger, Jeff. ”A lower bound for the smallest eigenvalue of the Laplacian.” Proceedings of the Princeton conference in honor of Professor S. Bochner. 1969.
2. Buser, Peter. ”A note on the isoperimetric constant.” Annales scientifiques de l’Ecole Normale Supérieure. Vol. 15. No. 2. 1982; Buser, Peter. ”Ueber den ersten Eigenwert des Laplace-Operators auf kompakten Flächen.” Commentarii Mathematici Helvetici 54.1 (1979): 477-493.
3. Alon, Noga, and Vitali D. Milman. ” λ_1 , isoperimetric inequalities for graphs, and super-concentrators.” Journal of Combinatorial Theory, Series B 38.1 (1985): 73-88; Alon, Noga. ”Eigenvalues and expanders.” Combinatorica 6.2 (1986): 83-96.
4. Fiedler, Miroslav. ”Algebraic connectivity of graphs.” Czechoslovak mathematical journal 23.2 (1973): 298-305.
5. Aldous, David. ”On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing.” Probability in the Engineering and Informational Sciences 1.01 (1987): 33-46.

6. Lawler, Gregory F., and Alan D. Sokal. "Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality." *Transactions of the American mathematical society* 309.2 (1988): 557-580.
7. Mihail, Milena. "Conductance and convergence of Markov chains-A combinatorial treatment of expanders." *IEEE Ann. Symp. on Foundations of computer science*, (1989).
8. Dodziuk, Jozef. "Difference equations, isoperimetric inequality and transience of certain random walks." *Transactions of the American Mathematical Society* 284.2 (1984): 787-794.
9. Chen, Fang, László Lovász, and Igor Pak. "Lifting Markov chains to speed up mixing." *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. ACM, 1999.
10. Aharonov, Dorit, et al. "Quantum walks on graphs." *Proceedings of the thirty-third annual ACM symposium on Theory of computing*. ACM, 2001.
11. Ford, Lester R., and Delbert R. Fulkerson. "Maximal flow through a network." *Canadian journal of Mathematics* 8.3 (1956): 399-404.
12. Mohseni, Masoud, et al. "Environment-assisted quantum walks in photosynthetic energy transfer." *The Journal of chemical physics* 129.17 (2008): 11B603.
13. John N. Tsitsiklis and Michael Athans (advisor). "Problems in decentralized decision making and computation." PhD Thesis, MIT (1984).
14. Murray, Richard and Reza Olfati Saber. "Consensus protocols for networks of dynamic agents." *Proc. American Control Conference* (2003).
15. Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of small-world networks." *nature* 393.6684 (1998): 440-442.
16. Jerrum, Mark, and Alistair Sinclair. "Approximating the permanent." *SIAM journal on computing* 18.6 (1989): 1149-1178.
17. Rezvaniyan, Alireza, Mohammad Rahmati, and Mohammad Reza Meybodi. "Sampling from complex networks using distributed learning automata." *Physica A: Statistical Mechanics and its Applications* 396 (2014): 224-234.
18. Diaconis, Persi, Susan Holmes, and Radford M. Neal. "Analysis of a nonreversible Markov chain sampler." *Annals of Applied Probability* (2000): 726-752.
19. Aaronson, Scott. "Quantum computing and hidden variables." *Physical Review A* 71.3 (2005): 032325.
20. Aldous, David, and Jim Fill. "Reversible Markov chains and random walks on graphs." (2002).
21. Levin, David Asher, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.
22. Apers, Simon, Alain Sarlette, and Francesco Ticozzi. "Lifting Markov chains to mix faster: Limits and Opportunities." In preparation for *IEEE Trans. Information Theory*, (2017).
23. Saloff-Coste, Laurent, and Jessica Zúniga. "Convergence of some time inhomogeneous Markov chains via spectral techniques." *Stochastic processes and their applications* 117.8 (2007): 961-979 ; Touri, Behrouz, and Angelia Nedić. "Alternative characterization of ergodicity for doubly stochastic chains." *Proc. IEEE Conf. on Decision and Control* (2011).
24. Mossel, Elchanan, Yuval Peres, and Alistair Sinclair. "Shuffling by semi-random transpositions." *Proc. IEEE Symp. on Foundations of Computer Science* (2004) ; Diaconis, Persi, and Arun Ram. *Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques*. Department of Statistics, Stanford University, (2000).
25. Hendrickx, Julien M., et al. "Graph diameter, eigenvalues, and minimum-time consensus." *Automatica* 50.2 (2014): 635-640 ; Hendrickx, Julien M., Guodong Shi, and Karl H. Johansson. "Finite-time consensus using stochastic matrices with positive diagonals." *IEEE Transactions on Automatic Control* 60.4 (2015): 1070-1073.
26. Hogben, Leslie. "Spectral graph theory and the inverse eigenvalue problem of a graph." *Electronic Journal of Linear Algebra* 14.1 (2005): 3.