

CNPq - Edital Universal

Modelagem probabilística da atividade cerebral

Antonio Galves

Proponente/Coordenador

Resumo

Dados experimentais sugerem que neurônios, sinapses e sistemas neuronais seguem um regime inerentemente estocástico. Surpreendentemente, enquanto fenômenos nos níveis micro e meso parecem apresentar uma aleatoriedade inerente, muitos fenômenos ao nível macro são fundamentalmente previsíveis. Para conciliar esta aparente contradição e alcançar uma formulação reveladora do fenômeno, o grupo se propõe utilizar uma abordagem probabilística inovadora e baseada na premissa de que a dinâmica neuronal pode ser descrita, em todas as escalas, por processos estocásticos com valores numa configuração espacial adequada. Desta forma pode-se fornecer uma abordagem uniforme para a consideração simultânea de diferentes escalas e conexões entre elas. Esta abordagem deve fornecer uma melhor compreensão da relação existente entre a atividade neuronal e medidas comportamentais.

A estratégia do grupo será utilizar o rigor matemático para abordar problemas centrais em neurociência: coleta e processamento de sinais, modelagem e aprendizado estatístico. Avanços nestas abordagens, além de ter grande impacto na neurociência, podem inaugurar uma nova área de pesquisa com interface entre teoria de probabilidade, combinatória, estatística e engenharia biomédica.

Modelos teóricos em neurociência necessitam de validação empírica. Por isso, o projeto contará com um grande banco de dados, coletados em diferentes experimentos por membros da equipe nos laboratórios associados.

1 Introdução

A rápida evolução da neurociência tem produzido uma grande massa de dados e uma série de fenômenos revelados por novos experimentos e intervenções. Entretanto, a compreensão teórica desses fenômenos não avançou com o mesmo ímpeto. Esta discrepância está descrita na página web do “Redwood Center for Theoretical Neuroscience” como um cenário rico em dados ainda que pobre teoricamente (*data-rich yet theory-poor*). Matemática e ciências correlatas são indispensáveis para que o equilíbrio seja alcançado.

Um primeiro desafio surge da complexidade das bases de dados. As ferramentas matemáticas e estatísticas atuais são insuficientes para tratar com bases de dados tão grandes e problemas de tão alta dimensionalidade. Neste cenário, novos procedimentos são necessários para obtenção de métodos estatísticos de alta precisão e computacionalmente viáveis.

Um segundo desafio surge da aparente aleatoriedade apresentada por dados neuronais. Uma publicação recente afirma (em tradução livre)

“Dados experimentais sugerem que neurônios, sinapses e sistemas neuronais são inerentemente estocásticos [...]. De fato, muitos estudos experimentais chegam a conclusão de que estímulos externos apenas modulam a alta atividade estocástica espontaneamente disparada pela atividade de redes corticais de neurônios [...]. Além disso, modelos tradicionais para computação neuronal são desafiados pelo fato de que sensores tradicionais de dados do ambiente são frequentemente ruidosos e ambíguos, exigindo que sistemas neuronais levem em conta a incerteza imposta por estímulos externos (Buesing *et al.*, 2011).”

Nos últimos anos a comunidade científica tem demonstrado um crescente interesse pelo desenvolvimento de modelos probabilísticos com base na atividade cerebral (Deco *et al.*, 2009; Harrison *et al.*, 2005; Toyozumi *et al.*, 2009; Cessac, 2011; Stevenson & Kording, 2011). Esta abordagem recente tem a vantagem de levar em consideração o conhecimento sobre o sistema neuronal para construir

modelos probabilísticos parsimoniosos. Entretanto, muitas das pesquisas probabilísticas sofrem de duas limitações que se opõem. Algumas pesquisas fornecem um tratamento matemático rigoroso de alguns modelos, mas sem confrontá-los com dados experimentais (veja, por exemplo, Cessac, 2011). Outras pesquisas utilizam modelos específicos para interpretar dados experimentais, porém sem um esforço sistemático para estudar propriedades matemáticas dos modelos envolvidos (Toyoizumi et al., 2009; Harrison et al., 2005), sendo analisados somente através de simulações numéricas.

Por outro lado, a teoria probabilística de sistemas com muitas componentes, inspirada na mecânica estatística clássica, foi proposta por inúmeros matemáticos nos anos 80. Provavelmente esta abordagem não apresentou influência em neurociência por ser baseada em ideias da mecânica estatística sem dados adequados para confrontar os modelos estocásticos multi-componentes propostos. Desde então uma grande massa de dados se tornou possível de coletar, graças a técnicas como eletroencefalografia de múltiplos canais (EEG) e ressonância magnética funcional (fMRI). A proposta do grupo é superar estas limitações através de uma conexão entre a teoria matemática e a análise estatística de dados experimentais.

2 Metodologia

A modelagem matemática dos fenômenos da atividade cerebral é tributária dos seguintes elementos:

1. Seu estudo é limitado pela insuficiente informação disponível atualmente relativa à conectividade entre neurônios e regiões do cérebro. Esta conectividade precisa ser estimada a partir de dados empíricos com registros de atividade individual de neurônios em uma população, ou de registros de atividade simultânea de sub-regiões do cérebro (Park & Terman, 2010).
2. Dados experimentais são extremamente sensíveis a ruídos e frequentemente levam a resultados ambíguos.

3. A atividade neural mostra uma grande variabilidade, com estímulos análogos provocando respostas diferentes em indivíduos diferentes, e mesmo em experiências repetidas com o mesmo indivíduo.
4. Dados neurais são coletados em diversas escalas que vão desde registros unitários até medidas mesoscópicas do cérebro e medidas comportamentais.

Esses pontos levam naturalmente a uma área específica da matemática que lida com fenômenos aleatórios resultantes da atividade de um grande número de componentes em diversas escalas, a saber, a área de processos e campos aleatórios. Essa área fornece o quadro geral de estudo, além de muitas ferramentas básicas já disponíveis. No entanto, as características específicas dos fenômenos neurais demandam a construção de novos objetos matemáticos. O presente projeto parte da idéia segundo a qual estados neurais podem ser descritos por medidas de probabilidade em espaços de configurações adequados, descrevendo a evolução de muitos agentes com interações de alcance variável no espaço e no tempo. Esta teoria deve combinar novos desenvolvimentos nas teorias de delineamento experimental, processamento de sinais, cadeias estocásticas, grafos aleatórios, mecânica estatística matemática e análise multivariada de dados. A equipe de matemáticos envolvidos no presente projeto é formada por especialistas nessas áreas.

O primeiro ingrediente da abordagem aqui proposta é uma nova teoria a ser desenvolvida de sistemas com muitas componentes e com interações de alcance variável no espaço e no tempo. Esses processos são extensões não triviais tanto dos sistemas markovianos de partículas introduzidos por Spitzer (1970), quanto das cadeias estocásticas de memória variável introduzidas por Rissanen (1983). Esses processos são intrinsecamente não estacionários no tempo e não homogêneos espacialmente e representam dinâmicas estocásticas que mudam em resposta tanto a estímulos exteriores constituídos por uma sucessão de amostras geradas por uma fonte aleatória, quanto em resposta a estímulos internos, através de um procedimento que evoca o de seleção estatística de modelos.

A teoria matemática desta nova classe de processos ainda está por ser construída. A equipe de matemáticos do projeto inclui especialistas em processos não-markovianos (Fernández & Galves, 2002; Collet *et al.*, 2005; Galves *et al.*, 2010b,a, 2012a), cadeias com memória de alcance variável (Galves & Leonardi, 2008; Collet *et al.*, 2008; Garivier & Leonardi, 2011) e campos e sistemas com interações de alcance variável (Galves *et al.*, 2012b; Cassandro *et al.*, 2012; Löcherbach & Orlandi, 2011). A equipe também tem uma grande experiência na modelagem de dados linguísticos e biológicos, com o foco principal no problema de seleção de modelos, caracterização e classificação (Leonardi, 2006; Cuesta-Albertos *et al.*, 2007; Busch *et al.*, 2009; Fraiman *et al.*, 2009; von Borries & Wang, 2009; Galves *et al.*, 2012a).

Na escala mesoscópica, os processos estocásticos descrevendo os fenômenos de atividades cerebrais assumem valores em grafos neurais sujeitos a interações (Bullmore & Bassett, 2011). A modelagem dessas evoluções temporais demandam um estudo rigoroso de grafos aleatórios e de suas propriedades. A equipe do projeto também inclui especialistas nesta área (Kohayakawa *et al.*, 2011, 2010b,a; Bollobás *et al.*, 2011; Alon *et al.*, 2011; Oliveira, 2010).

Delinear modelos adequados para os fenômenos descritos acima é uma tarefa desafiadora. Procedimentos de seleção de modelos devem ser embasados numa intensa atividade experimental. Enquanto modelos probabilísticos propostos acima são suficientemente gerais para fornecer uma base matemática sólida, a análise estatística dos dados e a seleção de modelos adequados é complexificada pela alta dimensionalidade dos modelos envolvidos. Somente a combinação da modelagem matemática com a validação experimental, confrontando predições e resultados empíricos pode levar a um novo nível de compreensão dos complexos fenômenos associados ao funcionamento do cérebro. Um tal empreendimento só pode ser levado a cabo por uma equipe multidisciplinar como aquela proposta aqui.

3 Objetivos

As principais aspirações do grupo em relação aos avanços na modelagem da atividade cerebral são:

1. Introduzir modelos probabilísticos com interações de alcance variável no espaço e no tempo que representem os sinais registrados por diferentes tecnologias, por exemplo eletroencefalografia (EEG) e ressonância magnética funcional (fMRI).
2. Propor critérios de estimação de parâmetros e seleção para estes novos modelos e para modelos relacionados já existentes na literatura
3. Estudar e caracterizar as distribuições de probabilidade sobre conjuntos de grafos subjacentes a diferentes estados neuronais.
4. Desenvolver teoria estatística não paramétrica para distribuições sobre grafos, como testes de aderência ou homogeneidade.
5. Aplicar os métodos desenvolvidos em dados reais para caracterizar e/ou classificar diferentes estados neurais a partir das características das respectivas distribuições de probabilidade.

4 Principais contribuições científicas

A abordagem probabilística, estatística e computacional proposta pelo grupo tem o potencial de conduzir a avanços significativos devido a três razões: *(i)* as lições aprendidas de tentativas anteriores mal sucedidas, *(ii)* a possibilidade de coleta de dados de EEG e de registros unitários *(iii)* a disponibilidade de novas técnicas matemáticas e estatísticas com suporte em uma sólida teoria probabilística. Membros do grupo são especialistas e até pioneiros nestas novas técnicas, como modelos de árvore de contexto, novas categorias de grafos aleatórios, novos procedimentos de seleção de modelos, agrupamento e classificação de dados superdimensionados. Além disso, o grupo terá condições de gerar grandes quantidades de dados

de EEG, possuindo especialistas no processamento e filtragem de sinais biopotenciais (sinais elétricos do corpo humano). A combinação de novas técnicas matemáticas e de delineamento experimental com o expertise dos pesquisadores do grupo deve contribuir para esta área de domínio.

5 Equipe

A equipe do projeto está integrada pelos seguintes pesquisadores.

5.1 Pesquisadores em instituições brasileiras

1. Antonio Galves (USP)
2. Cláudia Domingues Vargas (UFRJ)
3. Fátima Cristina Smith Erthal (UFRJ)
4. Florencia Leonardi (USP)
5. George Freitas von Borries (UNB)
6. Jesús E. Garcia (UNICAMP)
7. Nancy Lopes Garcia (UNICAMP)
8. Robert Morris (IMPA)
9. Roberto Imbuzeiro Oliveira (IMPA)
10. Sergio Neuenschwander (UFRN)
11. Sidarta Ribeiro (UFRN)
12. Yoshiharu Kohayakawa (USP)

5.2 Pesquisadores em instituições estrangeiras

1. Daniel Fraiman (Universidade de San Andrés)
2. Daniel Takahashi (Universidade de Princeton)
3. Eva Löcherbach (Universidade de Cergy-Pontoise)
4. Márzio Cassandro (Universidade de Roma “La Sapienza”)
5. Pierre Collet (CNRS/Escola Politécnica)
6. Ricardo Fraiman (Universidade de San Andrés)
7. Ricardo Freitas von Borries (Universidade de Texas em El Paso)
8. Roberto Fernández (Universidade de Utrecht)
9. Valeria Della Maggiore (Universidade de Buenos Aires)

5.3 Alunos de pós-graduação e pós-doutorandos

1. Alex Rodrigo dos Santos Souza (mestrado/UNICAMP)
2. Aline Duarte de Oliveira (doutorado/USP)
3. Andressa Cerqueira (mestrado/USP)
4. Bruno Monte de Castro (mestrado/USP)
5. Eduardo Furtado Martins (doutorado/UFRJ)
6. Estéfano Alves de Souza (doutorado/USP)
7. Guilherme Ost de Aguiar (doutorado/USP)
8. Iara Moreira Frondana (doutorado/USP)
9. Karina Yuriko Yaginuma (doutorado/USP)
10. Loyane Christina Soares Rocha (mestrado/UNB)

11. Sebastian Hofle (Bolsista DTI nível II do CNPq)
12. Simon Griffiths (pós-doutorado/IMPA)
13. Thaysa Guimarães Souza (mestrado/UNB)
14. Thiago Lemos (pós-doutorado/UFRJ)

6 Orçamento

6.1 Material permanente

Descrição	Preço total estimado
Servidores para grandes processamentos de dados	R\$ 25.000,00
Computadores portáteis	R\$ 35.000,00
Total	R\$ 60.000,00

6.2 Passagens e diárias para visitas técnicas

Descrição	Preço total estimado
Passagens aéreas nacionais e internacionais	R\$ 20.000,00
Diárias nacionais e internacionais	R\$ 40.000,00
Total	R\$ 60.000,00

7 Cronograma

Para o desenvolvimento do projeto estão previstas três etapas consecutivas com uma duração de 12 meses cada uma, a saber:

1ra Etapa: introdução de novos modelos com base nos fenômenos cerebrais. Estudo de propriedades teóricas de sistemas com interação de alcance variável, no espaço e no tempo.

2da Etapa: introdução de critérios de estimação de parâmetros e seleção de modelos para a nova classe de sistemas com interação de alcance variável

desenvolvida pelo grupo. Desenvolvimento de técnicas estatísticas como testes de hipóteses de aderência e homogeneidade.

3ra Etapa: modelagem com base nos resultados obtidos nas etapas anteriores de dados de eletroencefalografia, registros unitários de neurônios, ressonância magnética funcional e estimulação magnética transcraniana.

8 Cooperação científica em andamento

A equipe do projeto já vem trabalhando de forma conjunta na caracterização das propriedades dos sistemas estocásticos com interação de alcance variável e em métodos de seleção de modelos para esta nova classe de processos, como demonstram os trabalhos já publicados, submetidos ou em fase final de redação (Galves & Leonardi, 2008; Collet *et al.*, 2008; Cassandro *et al.*, 2012). Também existem trabalhos prévios da equipe sobre metodologia estatística para distribuições de árvores (uma classe particular de grafos) e sobre técnicas de agrupamento para dados superdimensionados (Busch *et al.*, 2009; von Borries & Wang, 2009).

Com relação às aplicações em neurociência, e fazendo parte das atividades de pesquisa do projeto MaCLinc (“*Matemática, Computação, Linguagem e Cérebro*”) da Universidade de São Paulo, existem diferentes linhas de cooperação com os laboratórios situados na UFRN e UFRJ, entre elas podemos citar:

- Descrição das atividades neuronais com base em processos de renovação: modelagem probabilística de registros unitários de atividade cerebral, estudo de propriedades teóricas e aplicações. Trabalho conjunto de Antonio Galves, Sidarta Ribeiro, Karina Yaginuma e Eva Löcherbach.
- Ritmos e seleção de modelos no cérebro humano: identificação dos padrões eletroencefalográficos produzidos pelo cérebro na identificação de diferentes ritmos, gerados por cadeias estocásticas de memória variável. Trabalho conjunto de Antonio Galves, Jesus E. Garcia, Sebastian Hoefle e Claudia D. Vargas.

- Construção de redes de interação baseadas na atividade eletroencefalográfica: reconstrução de redes de interações cerebrais e estudo das propriedades estatísticas. Trabalho em conjunto de Florencia Leonardi, Cláudia D. Vargas, Andressa Cerqueira e Daniel Fraiman.

9 Infra-estrutura e apoio técnico disponível

Na Universidade de São Paulo a equipe conta com a infra-estrutura do Instituto de Matemática e Estatística, onde está sediado o Núcleo de Apoio à Pesquisa NUMEC/MacLinc. O prédio conta com dois laboratórios de computação com um total de 35 computadores Mac mini e um laboratório de análise de dados (com dois servidores Mac Pro). Este prédio conta com o apoio técnico de uma secretária e dois bolsistas de treinamento técnico no nível de graduação.

As outras instituições envolvidas no presente projeto, e que oferecem o suporte de infra-estrutura e apoio técnico disponível, são a Universidade Estadual de Campinas (por meio do Instituto de Matemática, Estatística e Computação Científica), a Universidade de Brasília (por meio do Departamento de Estatística do Instituto de Ciências Exatas) e o Instituto de Matemática Pura e Aplicada (IMPA).

A equipe do projeto contará também com os recursos de infra-estrutura e apoio técnico dos laboratórios de neurociência associados. No laboratório de Neurobiologia II, localizado no Instituto de Biofísica Carlos Chagas Filho da Universidade federal do Rio de Janeiro (UFRJ), existem instaladas salas de experimentos para registros eletroencefalográficos, fisiológicos e comportamentais em humanos. Há montagens para registro do tempo de reação, e 4 sistemas de registro de sinais biológicos BIOPAC (MP100) para eletroencefalografia (3 canais), eletrooculografia, eletromiografia, eletrocardiografia, pletismografia e condutância da pele. Há também montagem para aquisição de sinal estabilométrico através de uma plataforma de força. Entre outros equipamentos dispõe-se de um eletroencefalógrafo de 32 canais, dois sistemas de estimulação magnética transcraniana

(TMS), Neurosoft e Magpro, e um sistema de EEG alta densidade (128 canais, GEODESIC), para a realização dos estudos psicofísicos e neurofisiológicos.

O Instituto do Cérebro, pertencente à Universidade Federal do Rio Grande do Norte (UFRN), localizado em Natal, conta com equipamentos de eletrofisiologia, imagem óptica assim como outros métodos de monitoramento da atividade cerebral.

O “Multi-Sensing-Processing and Learning Lab” (MSPL) está localizado no Departamento de Engenharia Elétrica e Computacional (ECE) da Universidade do Texas em El Paso (UTEP), Estados Unidos. O MSPL está equipado com instrumentação moderna para aquisição de potenciais bioelétricos e geração e aquisição de sinais de banda larga. Entre a instrumentação para aquisição de potenciais bioelétricos o laboratório inclui ActiveTwo (Biosemi) conversor analógico-digital (2 unidades com 128 canais cada, 24 bits de resolução e 2048 amostras por segundo), ActiveTwo conversor auxiliar analógico-digital para aquisição sincronizada (1 unidade com 8 canais, 24 bits de resolução e 2048 amostras por segundo), ActiveTwo eletrodo tipo pino com pré-amplificador para EEG (8 conjuntos de 32 eletrodos cada), ActiveTwo eletrodo tipo plano com pré-amplificador para EMG (2 conjuntos de 32 eletrodos cada), ActiveTwo eletrodo de carbono em fita com pré-amplificadores para ECG (4 conjuntos de 32 eletrodos cada), ActiveTwo para medida de condutividade da pele, MultiAmp gerador de sinais e calibrador para uso com bioamplificadores, e CardioSoft (GE) para aquisição de ECG (15 canais). O MSPL também está equipado com uma câmera de infravermelho de alta resolução (640x480 pixels com comprimento de onda de 7 a 14 microns). Além dessa instrumentação, o MSPL possui 4 computadores pessoais e um servidor de dados, todos conectados a Internet. O MSPL também dispõe de livre acesso ao grupo de computadores do “Distributed Computing Lab” (DCL) do ECE (computador “front end,” 21 nós de computação, e 2 nós de memória, cada um com 2 CPUs Intel quad-core num total de 48 CPUs).

10 Recursos de outras fontes

O projeto conta com importantes contrapartidas estaduais e de cooperação internacional, que podem complementar parte dos recursos necessários para o custeio das atividades de pesquisa. Entre elas, destacamos o projeto MaCLinC (“*Matemática, Computação, Linguagem e Cérebro*”), ao qual muitos dos pesquisadores desta equipe pertencem, tendo recebido da Universidade de São Paulo o valor de R\$1.998.022,00 para o período de Julho de 2011 a Junho de 2014. Outras contrapartidas importantes são o projeto de cooperação USP-COFECUB “*Stochastic systems with interaction of variable range*”, para o período de Março de 2010 a Fevereiro de 2014 (no valor de R\$ 62.661,28) e o programa CAPES-MathAmSud com o projeto “*Stochastic Structure of Large Interacting Systems*”, para o período de Fevereiro de 2011 a Janeiro de 2013 (tendo recebido R\$12.000,00). Além disso, três pesquisadores da equipe são bolsista de produtividade em pesquisa do CNPq no nível 1 (dois pesquisadores nível 1A e um pesquisador nível 1C).

Referências

- ALON, N., BALOGH, J., BOLLOBÁS, B. & MORRIS, R. (2011). The structure of almost all graphs in a hereditary property. *J. Combin. Theory Ser. B* **101**(2), 85–110.
- BOLLOBÁS, B., BRIGHTWELL, G. & MORRIS, R. (2011). Shadows of ordered graphs. *J. Combin. Theory Ser. A* **118**(3), 729–747.
- BUESING, L., BILL, J., NESSLER, B. & MAASS, W. (2011). Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput. Biol.* **7**(11), 1–22.
- BULLMORE, E. & BASSETT, D. (2011). Brain graphs: Graphical models of the human brain connectome. In: *Annual Review of Clinical Psychology* (NOLENHOEKSEMA, S., CANNON, T. & WIDIGER, T., eds.), vol. 7. Annual Reviews, pp. 113–140.

- BUSCH, J. R., FERRARI, P. A., FLESIA, A. G., FRAIMAN, R., GRYNBERG, S. P. & LEONARDI, F. (2009). Testing statistical hypothesis on random trees and applications to the protein classification problem. *Ann. Appl. Stat.* **3**(2), 542–563.
- CASSANDRO, M., GALVES, A. & LÖCHERBACH, E. (2012). Partially observed markov random fields are variable neighborhood random fields. *Journal of Statistical Physics* **147**, 795–807.
- CESSAC, B. (2011). Statistics of spike trains in conductance-based neural networks: Rigorous results. *J. Math. Neurosci.* **1**(8), 1–42.
- COLLET, P., DUARTE, D. & GALVES, A. (2005). Bootstrap central limit theorem for chains of infinite order via Markov approximations. *Markov Process. Related Fields* **11**(3), 443–464.
- COLLET, P., GALVES, A. & LEONARDI, F. (2008). Random perturbations of stochastic processes with unbounded variable length memory. *Electron. J. Probab.* **13**, no. 48, 1345–1361.
- CUESTA-ALBERTOS, J., FRAIMAN, R., GALVES, A., GARCIA, J. & SVARC, M. (2007). Identifying rhythmic classes of languages using their sonority: a Kolmogorov-Smirnov approach. *J. Appl. Stat.* **34**, 749–761.
- DECO, G., ROLLS, E. & ROMO, R. (2009). Stochastic dynamics as a principle of brain function. *Prog. Neurobiol.* **88**(1), 1–16.
- FERNÁNDEZ, R. & GALVES, A. (2002). Markov approximations of chains of infinite order. *Bull. Braz. Math. Soc. (N.S.)* **33**(3), 295–306. Fifth Brazilian School in Probability (Ubatuba, 2001).
- FRAIMAN, D., BALENZUELA, P., FOSS, J. & CHIALVO, D. R. (2009). Ising-like dynamics in large-scale functional brain networks. *PHYSICAL REVIEW E* **79**(6).

- GALVES, A., GALVES, C., GARCIA, J. E., GARCIA, N. L. & LEONARDI, F. (2012a). Context tree selection and linguistic rhythm retrieval from written texts. *Ann. Appl. Stat.* **6**(1), 186–209.
- GALVES, A., GARCIA, N., LÖCHERBACH, E. & ORLANDI, E. (2012b). Kalikow-type decomposition for multicolor infinite range particle systems. *Annals of Applied Probability (accepted)* .
- GALVES, A., GARCIA, N. L. & PRIEUR, C. (2010a). Perfect simulation of a coupling achieving the \bar{d} -distance between ordered pairs of binary chains of infinite order. *J. Stat. Phys.* **141**(4), 669–682.
- GALVES, A. & LEONARDI, F. (2008). *Exponential inequalities for empirical unbounded context trees*, vol. 60 of *Progress in Probability*. Birkhauser, pp. 257–270.
- GALVES, A., LÖCHERBACH, E. & ORLANDI, E. (2010b). Perfect simulation of infinite range Gibbs Measures and coupling with their finite range approximations. *J. Stat. Phys.* **138**, 476–495.
- GARIVIER, A. & LEONARDI, F. (2011). Context tree selection: a unifying view. *Stochastic Process. Appl.* **121**(11), 2488–2506.
- HARRISON, L., DAVID, O. & FRISTON, K. (2005). Stochastic models of neuronal dynamics. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**(1457), 1075–1091.
- KOHAYAKAWA, Y., NAGLE, B., RÖDL, V. & SCHACHT, M. (2010a). Weak hypergraph regularity and linear hypergraphs. *J. Combin. Theory Ser. B* **100**(2), 151–160.
- KOHAYAKAWA, Y., RÖDL, V., SCHACHT, M. & SKOKAN, J. (2010b). On the triangle removal lemma for subgraphs of sparse pseudorandom graphs. In: *An irregular mind*, vol. 21 of *Bolyai Soc. Math. Stud.* Budapest: János Bolyai Math. Soc., pp. 359–404.

- KOHAYAKAWA, Y., RÖDL, V., SCHACHT, M. & SZEMERÉDI, E. (2011). Sparse partition universal graphs for graphs of bounded degree. *Adv. Math.* **226**(6), 5041–5065.
- LEONARDI, F. (2006). A generalization of the PST algorithm: modeling the sparse nature of protein sequences. *Bioinformatics* **22**(11), 1302–1307.
- LÖCHERBACH, E. & ORLANDI, E. (2011). Neighborhood radius estimation for variable-neighborhood random fields. *Stochastic Process. Appl.* **121**(9), 2151–2185.
- OLIVEIRA, R. I. (2010). The spectrum of random k -lifts of large graphs (with possibly large k). *J. Comb.* **1**(3-4), 285–306.
- PARK, C. & TERMAN, D. (2010). Irregular behavior in an excitatory-inhibitory neuronal network. *Chaos* **20**(2), 1054–1500.
- RISSANEN, J. (1983). A universal data compression system. *IEEE Trans. Inform. Theory* **29**(5), 656–664.
- SPITZER, F. (1970). Interaction of Markov Processes. *Adv. Math.* **5**, 246–290.
- STEVENSON, I. & KORDING, K. (2011). How advances in neural recording affect data analysis. *Nat. Publ. Group* **14**(2), 139–142.
- TOYOIZUMI, T., RAD, K. & PANINSKI, L. (2009). Mean-field approximations for coupled populations of generalized linear model spiking neurons with markov refractoriness. *Neural Comput.* **21**(5), 1203–1243.
- VON BORRIES, G. & WANG, H. (2009). Partition clustering of high dimensional low sample size data based on p-values. *Computational Statistics and Data Analysis* **53**, 3987–3998.