Two part envelopes for rejection sampling of some completely random measures

Jim Griffin[a]

[a] *University College London*

Abstract

This paper shows that rejection sampling with a two-piece Lévy intensity envelope can outperform both the Ferguson-Klass algorithm and previously proposed envelopes for simulating realisations of completely random measures typically used in Bayesian nonparametric statistics.

Keywords: Bayesian nonparametrics; gamma process; beta process; generalized gamma process; stable-beta process; Ferguson-Klass algorithm

1. Introduction

Completely random measures (CRMs) play a central role in Bayesian nonparametric statistics. Most prominently, CRMs have been used to define tractable priors for survival analysis (Doksum, 1974) and their normalisations have been used to define tractable priors for random distributions in density estimation building on the initial work of Regazzini et al. (2003). A full description of posterior inference in this class of models is given by James et al. (2009) and a review of work in this area is provided by Lijoi and Prünster (2010). Caron and Fox (2017) and Todeschini and Caron (2016) discussed the use of gamma and generalized gamma processes to define models for sparse graphs. Thibaux and Jordan (2007) showed that the Indian buffet process (Ghahramani and Griffiths, 2006), which is a prior for an infinite feature model, can be defined using a beta process Hjort (1990), which was extended to the stable-beta process by Teh and Gorur (2009).

CRMs without fixed points of discontinuity (Kingman, 1967) can be represented in the form

$$G = \sum_{k=1}^{\infty} J_k \delta_{\theta_k}$$

where δ_x is the Dirac delta function and $(J_1, \theta_1), (J_2, \theta_2), (J_3, \theta_3), \ldots$ are a realisation of a Lévy intensity $\bar{\nu}(dJ, d\theta)$. In this work, it is assumed that $\bar{\nu}(dJ, d\theta) = \nu(dJ)\alpha(d\theta)$, where α is a probability measure. Often, $J_1, J_2, J_3, \ldots$ are referred to as jump sizes, $\theta_1, \theta_2, \theta_3, \ldots$ are referred to as jump locations and I will refer to ν as the jump intensity. The jump intensities of some popular examples are given in Table 1, where all processes, apart from the σ-stable, are standardized so that $\mathbb{E}[\sum_{i=1}^{\infty} J_i] = M$.

Simulating realisations of these processes is an important problem. Some MCMC algorithms for normalized random measures involve simulation of these processes or tilted versions of these measures (Griffin and Walker, 2011; Favaro and Teh, 2013). The Gibbs sampler for the sparse graph model in Todeschini and Caron (2016) uses an approximation of the sum of all jumps of a CRM. More generally, there may be interest in simulating these measures even in situations where the CRM is integrated from the model in the MCMC. The most widely-used method is due to Ferguson and Klass (1972) but other methods have been developed and are comprehensively reviewed in Campbell et al. (2018) in the context of Bayesian nonparametrics.

Email address: `j.griffin@ucl.ac.uk` (Jim Griffin)

Process	$\nu(x)$	Parameter range
Gamma	$Mx^{-1}\exp\{-x\}$	$x>0,\ M>0$
σ-stable	$\frac{\sigma}{\Gamma(1-\sigma)}x^{-1-\sigma}$	$x>0,\ 0<\sigma<1$
Beta	$M\,cx^{-1}(1-x)^{c-1}$	$0<x<1,\ c>0$
Generalized gamma	$M\frac{a^{1-\sigma}}{\Gamma(1-\sigma)}x^{-1-\sigma}\exp\{-ax\}$	$x>0,\ M>0,\ 0<\sigma<1,\ a>0$
Stable-Beta	$M\frac{\Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma)}x^{-1-\sigma}(1-x)^{c+\sigma-1}$	$0<x<1,\ 0<\sigma<1,\ c>0$

Table 1: The jump intensity for various CRM's

In this paper, I will concentrate on the class of rejection sampling methods described in Rosiński (2001), where a realisation of a CRM is generated by thinning a realisation from an enveloping CRM. I will discuss a simple construction of the enveloping CRM which can outperform previous suggestions in the literature for processes used in Bayesian nonparametrics. The rest of the paper is organized as follows. Section 2 describes the Ferguson-Klass and rejection algorithms for non-Gaussian Lévy processes, and proposes a two-piece construction of the jump intensity of the enveloping CRM. Section 4 illustrates the construction for some processes used in Bayesian nonparametrics. Section 5 reports on the performance of these algorithms compared to some previous suggestions. Section 6 provides a brief discussion.

2. Ferguson-Klass algorithm, rejection algorithms and two-piece Lévy envelopes

The most widely used method for simulating a non-Gaussian Lévy process is due to Ferguson and Klass (1972). Suppose that we wish to simulate a realisation of a non-Gaussian Lévy process, G, with Lévy intensity $\bar{\nu}(dJ, d\theta) = \nu(dJ)\,\alpha(d\theta)$. Define the tail mass function η_ν of a jump intensity ν to be $\eta_\nu(x) = \int_x^\infty \nu(z)\,dz$. Suppose that $E_1, E_2, \dots$ are the points of a Poisson process with intensity 1. Ferguson and Klass (1972) showed that G can be represented as

$$G = \sum_{k=1}^{\infty} J_k \delta_{\theta_k}$$

where $J_k = \eta_\nu^{-1}(E_j)$ and $\theta_1, \theta_2, \theta_3, \dots \overset{i.i.d.}{\sim} \alpha$. This construction implies that the J_k's are monotonically decreasing. In practice, this algorithm can be slow to run since the function η_ν will need to be numerically integrated and inverted for many processes. In the five processes in Table 1, the tail mass function is only available analytically for the σ-stable process (although, the tail mass function of the gamma process is the exponential-integral function for which a continued fraction method is available).

An alternative method is rejection sampling (Rosiński, 2001). Suppose, again, that we want to simulate a realisation of G and, furthermore, that we can define an enveloping Lévy process $\tilde{G}$ with Lévy intensity $\bar{\phi}(dJ, d\theta) = \phi(dJ)\,\alpha(d\theta)$ for which $\nu(x) \leq \phi(x)$ for all x, and for which η_ϕ can be analytically inverted. The algorithm works by first simulating a realisation $(\tilde{J}_1, \tilde{\theta}_1), (\tilde{J}_2, \tilde{\theta}_2), (\tilde{J}_3, \tilde{\theta}_3), \dots$ of $\tilde{G}$ using the Ferguson-Klass algorithm and then simulating selection variables $S_1, S_2, S_3, \dots$ which are independent Bernoulli random variables with success probability $\frac{\nu(\tilde{J}_k)}{\phi(\tilde{J}_k)}$. Then

$$G = \sum_{k=1}^{\infty} S_k \tilde{J}_k \delta_{\tilde{\theta}_k}.$$

The number of rejected points, $\sum_{k=1}^{\infty} \mathrm{I}(S_k = 0)$, is Poisson distributed with mean $\int_0^\infty [\phi(x) - \nu(x)]\,dx$ (Campbell et al., 2018). This integral will become smaller as $\phi(x)$ becomes closer to $\nu(x)$. The use of the Ferguson-Klass algorithm to simulate $\tilde{G}$ implies that the jumps of G are monotonically decreasing.

A simple method for constructing the jump intensity of $\tilde{G}$ uses a *two-piece jump intensity*. Suppose that we can write $\nu(x) = \nu_1(x)\,\nu_2(x)$ where $\nu_2(x) \leq 1$ for all x, $\nu_1(x)$ is non-increasing and the tail mass integral of $\nu_1(x)$ and $\nu_2(x)$ are analytically invertible. A suitable enveloping jump intensity is

$$\phi(x) = \begin{cases} \nu_1(x) & x < b \\ \nu_1(b)\,\nu_2(x) & x \geq b \end{cases}$$

which is well-defined for all b (for a fixed b, we need $\nu_2(x) \leq 1$ for $x < b$ and $\nu_1(x) \leq \nu_1(b)$ for $x \geq b$). The parameter b can be chosen to minimize the average number of rejections, which is denoted b_{opt}. An alternative adaptive thinning method for a CRM truncated to $[S, \infty]$ has been studied by Favaro and Teh (2013). In the following section, the two-piece construction is illustrated for the most popular classes of CRMs in Bayesian nonparametrics and b_{opt} is discussed.

3. Two-piece Lévy intensity envelopes for some non-Gaussian Lévy processes

3.1. Gamma process

Rosiński (2001) suggested an enveloping jump intensity

$$\phi(x) = M\,x^{-1}(1+x)^{-1}$$

which was termed the Lomax process by Campbell et al. (2018). The natural two-piece jump intensity is

$$\phi(x) = \begin{cases} M\,x^{-1} & x < b \\ M\,b^{-1}\exp\{-x\} & x \geq b \end{cases}.$$

This implies that

$$\eta_\phi^{-1}(x) = \begin{cases} b\,\exp\{b^{-1}\exp\{-b\} - \frac{x}{M}\} & x > M\,b^{-1}\exp\{-b\} \\ -\log(b\frac{x}{M}) & x \leq M\,b^{-1}\exp\{-b\} \end{cases}.$$

The optimal value b_{opt} solves $b_{opt} - b_{opt}e^{-b_{opt}} - e^{-b_{opt}} = 0$ which is approximately equal to 0.8065.

3.2. Generalized gamma process

The gamma process can be generalized by introducing a tilting parameter σ, which leads to a heavy tailed distribution for $\sum_{i=1}^{\infty} J_i$. Generalized gamma processes are generally considered to be more computationally demanding to simulate than gamma processes. It is useful to first use the re-parameterization $z = ax$ leading to the jump intensity

$$\nu(z) = \frac{M\,a}{\Gamma(1-\sigma)} z^{-1-\sigma}\exp\{-z\}.$$

If we generate jumps $K_1, K_2, K_3, \ldots$ from this jump intensity then $J_i = \frac{K_i}{a}$ will be the jump of the original generalized gamma process. A suitable two-piece enveloping jump intensity is

$$\phi(x) = \begin{cases} \frac{M\,a}{\Gamma(1-\sigma)}\,x^{-1-\sigma} & z < b \\ \frac{M\,a}{\Gamma(1-\sigma)}\,b^{-1-\sigma}\exp\{-z\} & z \geq b \end{cases}.$$

This choice of envelope implies that

$$\eta_\phi^{-1}(x) = \begin{cases} [\sigma\frac{\Gamma(1-\sigma)}{a}\frac{x}{M} + b^{-\sigma} - \sigma b^{-1-\sigma}\exp\{-b\})]^{-1/\sigma} & x > \frac{M\,a}{\Gamma(1-\sigma)}b^{-1-\sigma}\exp\{-b\} \\ -\log(b^{\sigma+1}\frac{\Gamma(1-\sigma)}{a}\frac{x}{M}) & x \leq \frac{M\,a}{\Gamma(1-\sigma)}b^{-1-\sigma}\exp\{-b\} \end{cases}.$$

The value of σ has a small effect on the value of b_{opt} and I suggest using $b_{opt} = 0.8065$. Campbell et al. (2018) suggest an envelope which corresponds to $b = \infty$.

3.3. Beta process

A suitable two-piece jump intensity envelope for the beta process is

$$\phi(x) = \begin{cases} M\, cx^{-1} & x < b \\ M\, cb^{-1}(1-x)^{c-1} & x \geq b \end{cases}.$$

This choice of envelope implies that

$$\eta_\phi^{-1}(x) = \begin{cases} b\, \exp\{-c^{-1}(\frac{x}{M} - b^{-1}(1-b)^c)\} & x > M\, b^{-1}(1-b)^c \\ 1 - (\frac{x}{M} b)^{1/c} & x \leq M\, b^{-1}(1-b)^c \end{cases}.$$

The value of b_{opt} as a function of c is shown in the first column of Figure 1, and $b_{opt} = \frac{4}{5c}$ is an extremely good approximation. Campbell et al. (2018) suggest an envelope which corresponds to $b = 1$.

3.4. Stable-beta process

A possible two-piece jump intensity envelope for the stable-beta process is

$$\phi(x) = \begin{cases} \frac{M\, \Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma)} x^{-1-\sigma} & x < b \\ \frac{M\, \Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma)} b^{-1-\sigma}\, (1-x)^{c+\sigma-1} & x \geq b \end{cases}.$$

This choice of jump intensity envelope implies that

$$\eta_\phi^{-1}(x) = \begin{cases} [\frac{\sigma\Gamma(1-\sigma)\Gamma(c+\sigma)}{\Gamma(1+c)} \frac{x}{M} - \frac{\sigma}{c+\sigma} b^{-1-\sigma}(1-b)^{c+\sigma} + b^{-\sigma}]^{-1/\sigma} & x > \frac{M\, \Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma+1)} b^{-1-\sigma}\, (1-b)^{c+\sigma} \\ 1 - (\frac{\Gamma(1-\sigma)\Gamma(c+\sigma+1)}{\Gamma(1+c)} b^{1+\sigma}\, \frac{x}{M})^{1/(c+\sigma)} & x \leq \frac{M\, \Gamma(1+c)}{\Gamma(1-\sigma)\Gamma(c+\sigma+1)} b^{-1-\sigma}\, (1-b)^{c+\sigma} \end{cases}.$$

The parameter σ has only a small effect on b_{opt} (see Figure 1) and a simple choice is $b = \frac{4}{5c}$, as for the beta process. Again, Campbell et al. (2018) suggest an envelope which corresponds to $b = 1$.

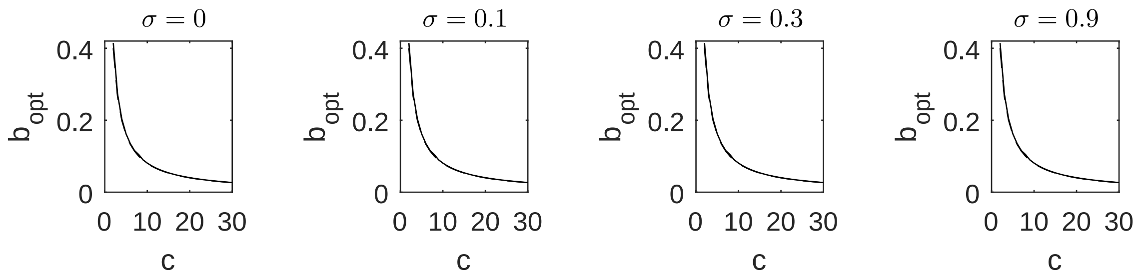

Figure 1: The value of b_{opt} as a function of c for the stable-beta process (solid line) with different values of σ and the function $b_{opt} = \frac{4}{5c}$ (dashed line).

4. Performance comparisons

The performance of the rejection algorithms with two-piece jump intensity was compared to direct application of the Ferguson-Klass (FK) algorithm and the previously suggested jump intensity envelopes of Rosiński (2001) (Lomax) and Campbell et al. (2018) (CHHB). The gamma, generalized gamma, beta and stable beta process were considered with a range of values for the parameters of these processes. In both FK and rejection algorithms, the 100 largest jumps of G were generated. Therefore, each algorithm produces values with the same truncation error but the rejection algorithms will tend to simulate more than 100 jumps from the enveloping CRM. This contrasts with the simulation study of Campbell et al. (2018) where a fixed number of jumps from $\tilde{G}$ are simulated and the truncation errors are compared. Arbel and Prünster

(2016) provide a much deeper discussion about truncation errors for CRMs. Each algorithm was run 10000 times to provide Monte Carlo estimates of the average number of rejections and the computational time. All code was written in Matlab. Numerical integrals were calculated using the built-in `integral` function and numerical inverses were calculated using the built-in `fzero` function. The average number of rejections from each envelope is linear in M and the results are shown in the on-line appendix.

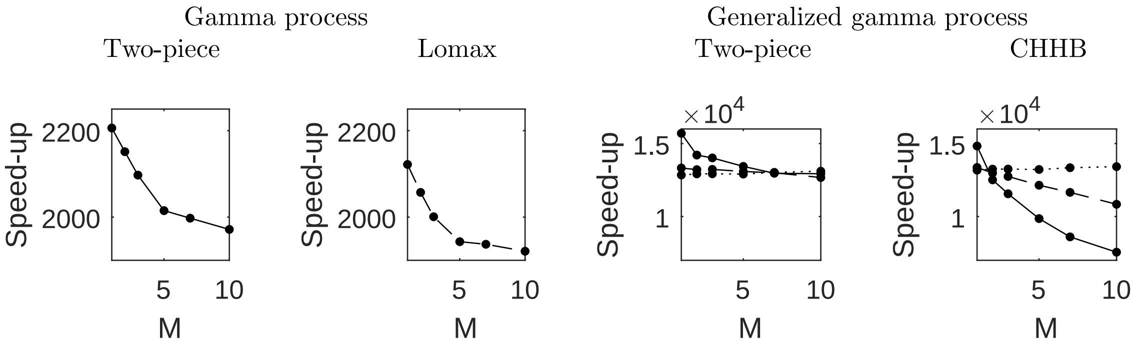

Figure 2: The speed-ups over the FK algorithm of the two-piece envelop and Lomax or CHHB envelope for the gamma and generalized gamma proceses. The value of σ are $\sigma = 0.1$ (solid line), $\sigma = 0.3$ (dashed line) and $\sigma = 0.9$ (dotted line).

Figure 2 shows speed-ups of the rejection algorithms over the FK algorithm for the gamma and generalized gamma processes. Both rejection algorithms are substantially faster than the FK algorithm for the gamma process (about 2000 times) and the generalized gamma process (between 8000 and 16000 times). The difference in speed-up is due to the availability of a continued fraction method for the tail-mass function of the gamma process but not the generalized gamma process, where numerical integration is required. The CHHB rejection envelope and the two-piece envelope have a similar speed-up for large σ but the difference increases as σ decreases. This reflect the number of rejections generated by each envelope. For example, when $\sigma = 0.1$ and $M = 10$, the CHHB rejection envelope has an average number of rejections of 100 whereas the two-piece envelope has an average number of rejections of 9. For the gamma process, the Lomax envelope gives a lower average number of rejections but the two-piece envelope is slightly faster due to the operations involved in the computation.

The results for beta and stable-beta processes are shown in Figure 3. The speed-up is substantial for both processes with a larger speed-up for smaller values of c. The additional speed-up of the two-piece envelope over the CHHB envelope increases with c. For example, with $c = 20$ and $M = 10$, the two-piece rejection envelope takes roughly half the time. This reflects that the difference in the number of rejections is increasing with c. For example, when $c = 20$ and $M = 10$, the two-piece envelope has an average of 20 rejections whereas the CHHB envelope has an average of 600 rejections for the beta process.

5. Discussion

This paper develops two-piece jump intensity envelopes for the most popular processes in Bayesian nonparametric modelling. These are shown to be several orders of magnitude faster than the popular Ferguson-Klass algorithm. This shows the potential of rejection sampling algorithms for efficient computation in Bayesian nonparametric analysis. The two-piece approach is also shown to outperform envelopes in Rosiński (2001) and Campbell et al. (2018) for some parameter values. This paper has concentrated on simple processes but there are other challenging problems such as simulating tilted versions of these processes.

References

Arbel, J. and I. Prünster (2016). A moment-matching Ferguson & Klass algorithm. *Statistics and Computing 27*, 3–17.

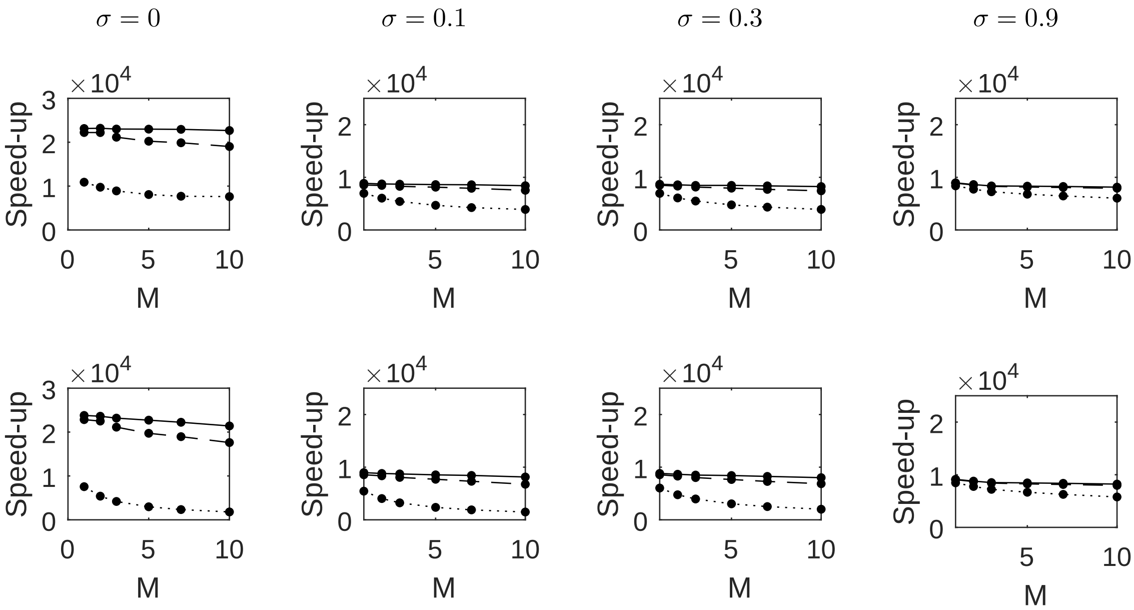

Figure 3: The speed-ups over the FK algorithm of the two-piece envelop (top row) and CHHB envelope (bottom row) for the stable-beta process. The value of c are $c = 2$ (solid line), $c = 3$ (dashed line) and $c = 20$ (dotted line).

Campbell, T., J. H. Huggins, J. P. How, and T. Broderick (2018). Truncated random measures. *Bernoulli*, to appear.

Caron, F. and E. B. Fox (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society, Series B 79*, 1295–1366.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability 2*, 183–201.

Favaro, S. and Y. W. Teh (2013). MCMC for normalized random measure mixture models. *Statistical Science 28*, 335–359.

Ferguson, T. S. and M. J. Klass (1972). A representation of independent increment processes without Gaussian components. *The Annals of Mathematical Statistics 43*, 1634–1643.

Ghahramani, Z. and T. L. Griffiths (2006). Infinite latent feature models and the Indian buffet process. In Y. Weiss, B. Schölkopf, and J. C. Platt (Eds.), *Advances in Neural Information Processing Systems 18*, pp. 475–482. MIT Press.

Griffin, J. E. and S. G. Walker (2011). Posterior simulation of normalized random measure mixtures. *Journal of Computational and Graphical Statistics 20*, 241–259.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for lifetime data. *The Annals of Statistics 18*, 1259–1294.

James, L. F., A. Lijoi, and I. Prünster (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics 36*, 76–97.

Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics 21*, 59–78.

Lijoi, A. and I. Prünster (2010). Models beyond the Dirichlet Process. In N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker (Eds.), *Bayesian nonparametrics*, pp. 80–136. Cambridge University Press.

Regazzini, E., A. Lijoi, and I. Prünster (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics 31*, 560–585.

Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In O. Barndorff-Nielson, S. Resnick, and T. Mikosch (Eds.), *Lévy processes: theory and applications*, Chapter VI, pp. 401–415. Boston: Bikhäuser.

Teh, Y. W. and D. Gorur (2009). Indian buffet processes with power-law behavior. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22*, pp. 1838–1846. Curran Associates, Inc.

Thibaux, R. and M. I. Jordan (2007, 21–24 Mar). Hierarchical beta processes and the Indian buffet process. In M. Meila and X. Shen (Eds.), *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, Volume 2 of *Proceedings of Machine Learning Research*, San Juan, Puerto Rico, pp. 564–571. PMLR.

Todeschini, A. and F. Caron (2016). Exchangeable random measures for sparse and modular graphs with overlapping communities. arXiv: 1602.0211.