Vertex cover on other graph ensembles: the effect of degree–degree correlations

Matti Peltomäki
Laboratory of Physics, Helsinki University of Technology, P.O. Box 1100, 02015 HUT, Espoo, Finland

I. INTRODUCTION

The vertex cover is a subset $V_{vc} \subset V$ of vertices of a graph $G(V, E)$ such that it contains at least one endpoint of each vertex $e \in E$. Solving for the minimum vertex cover, i.e. one with the smallest number of vertices, is in general an NP-complete problem. However, in practice many relatively simple algorithms are able to produce optimum or nearly-optimum results with short running times for some instances of the graphs. In particular, the time complexity depends heavily on the fraction of covered nodes for some algorithms [1] and thus on the structure of the graph. A big part of the literature deals with the issue on the traditional random graph ensemble $G(N, c/N)$.

In this summary, I go through the results of Ref. [2] by Vázquez and Weigt, where the minimum vertex-cover problem is studied on general random networks with arbitrary degree distribution and degree–degree correlations. In particular, the correlations are made tunable, and their effect on the minimum vertex-cover size and the replica symmetry and breaking thereof of the system are discussed. In the end, the results are numerically applied to power-law random graphs with positive correlations. These, up to some extent, resemble observations on real-world graphs [3, 4], and solving for the minimum vertex cover on them has some applications in, for instance, network traffic monitoring [5].

II. CORRELATED GENERALIZED RANDOM GRAPHS

Correlated generalized random graphs are a generalization of the Erdös-Renyi random graph ensemble (see, for instance, [6]). The degree distribution can be arbitrarily chosen and given degree–degree correlations can be imposed. To arrive at an ensemble of generalized random graphs, consider a set of undirected graphs with N vertices and an arbitrary degree distribution p_d. If one follows a randomly chosen edge, a vertex of degree $d + 1$ is recovered with probability

$$q_d = \frac{(d+1)p_{d+1}}{\langle d \rangle}\,, \tag{1}$$

where $\langle d \rangle$ is the average degree. The number of additional edges is called the *excess degree.*

Correlations between the degrees of adjacent vertices are incorporated as follows. In an uncorrelated graph, the probability that a randomly selected edge connects two vertices of excess degrees d and d' is $(2 - \delta_{d,d,'})q_d q_{d'}$. The prefactor $2 - \delta_{d,d'}$ comes from the fact that if d and d' differ, vertices of degrees d and d' can be found in two ways, first one with degree d and the latter with degree d' or vice versa. If $d = d'$, a similar effect does not take place. In correlated nets, the above probability is generalized to the following form

$$(2 - \delta_{d,d'})e_{dd'}\,, \tag{2}$$

where $e_{dd'}$ is related to the conditional probability $P(d|d')$ that a vertex of excess degree d is arrived at following any edge emanating from a vertex of excess degree d'

$$e_{dd'} = q_{d'} P(d|d')\,. \tag{3}$$

In this summary, the form

$$q_{dd'} = q_d(r\delta_{d,d'} + (1-r)q_{d'}) \tag{4}$$

is chosen, following [2], to facilitate easy generation of the graphs for testing purposes.

Such graphs can be generated using the so-called configuration model [3, 7]. For each node i draw a random degree d_i from the probability distribution p_d, restricted such that the sum of the degrees is even. Now create a set S of *stubs* (or half-edges) where each node appears with multiplicity d_i. The cardinality of this set is now $|S| = 2m$, where m is the number of edges. For each edge, select first a random stub with uniform probability. With probability r pair the previously selected stub with a randomly selected one with the same degree. Otherwise, with probability $1 - r$ pair the previously selected stub with a random one. The details of this procedure might affect the correlation properties of the resulting graph, see Sec. V.

III. GENERAL LATTICE GAS ON GRAPHS

Consider an undirected graph with adjacency matrix J_{ij}. A general lattice gas on such graph is defined by the Hamiltonian

$$-\beta H = \sum_{i<j} J_{ij} w(x_i, x_j) + \mu \sum_i x_i \,, \tag{5}$$

where the microscopic degrees of freedom take two values, $x_i = 0, 1$. The value $x_i = 1$ is associated with a particle lying on site i. The chemical potential is denoted by μ and the function $w(x_i, x_j)$ describes the interaction between the particles. The ferromagnetic Ising model can be recovered by choosing

$$w(x_i, x_j) = (2x_i - 1)(2x_j - 1) \,. \tag{6}$$

In Ref. [2], Vazquez and Weigt perform a cavity calculation of the system, which they also apply to the minimum vertex cover. In this summary, I will next go through the main parts of the calculation.

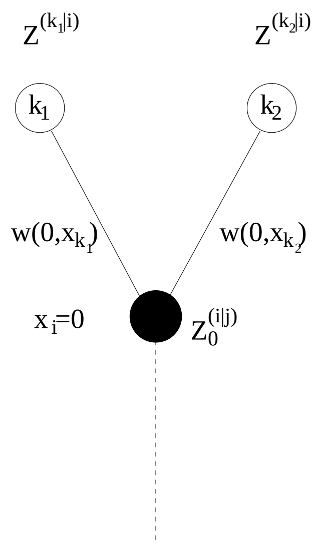

FIG. 1: Notation for and illustration of the partition function calculation of Eqs. (7). Edge (i, j) has been removed and the index k runs over all other neighbours of i except j. Eq. (8) is obtained similarly but by setting $x_i = 1$.

Consider an undirected graph, and assume that it is locally treelike. Choose an arbitrary edge (i, j) and remove it. Let index k run over all other neighbours of i except j and consider the subtree rooted in i. The notation is illustrated in Fig. 1. The partition functions for this subtree can be written recursively as follows. Restricting the values of x_i to zero, the partition function can be written as

$$Z_0^{(i|j)} = \prod_{k \neq j | J_{ik}=1} \left(e^{w(0,0)} Z_0^{(k|i)} + e^{w(0,1)} Z_1^{(k|i)} \right) , \tag{7}$$

where the restricted partition functions of the k-rooted subtrees can be considered independent since when the edges (i, k) are removed, all k-rooted subtrees are independent because of the assumption that the graph is locally treelike. Similarly, in the case of $x_i = 1$ one gets

$$Z_1^{(i|j)} = e^{\mu} \prod_{k \neq j | J_{ik}=1} \left(e^{w(1,0)} Z_0^{(k|i)} + e^{w(1,1)} Z_1^{(k|i)} \right) , \tag{8}$$

where now the contribution e^{μ} comes from the presence of a particle at site i.

The effective fields are defined as

$$h_{(i|j)} = \ln \frac{Z_1^{(i|j)}}{Z_0^{(i|j)}} \,. \tag{9}$$

To see the physical meaning of this quantity, consider an isolated particle in an external field described by the Hamiltonian $-\beta H = hx$. For this particle, the restricted partition functions read $Z_0 = e^0 = 1$ and $Z_1 = e^h$, from which the field can be calculated as $h = \ln \frac{Z_1}{Z_0}$. In this light, the field $h_{(i|j)}$ of Eq. (9) can be regarded as a generalization of an effective local field for the general lattice gas. Using Eqs. (7) and (8), the effective fields obey the recursion relation

$$h_{(i|j)} = \mu + \sum_{k \neq j | J_{ik}=1} u(h_{k|i}) \,, \tag{10}$$

where the auxiliary function u is

$$u(h_{k|i}) = \ln\left(\frac{e^{w(1,0)} + e^{w(1,1)+h_{(k|i)}}}{e^{w(0,0)} + e^{w(0,1)+h_{(k|i)}}}\right) \tag{11}$$

Eq. (10) can be considered as fixed point iteration: given the fields for all vertices, one can substitute them on the right-hand side, evaluate Eq. (10) and arrive at the iterated effective field for all sites. Roughly speaking, the iteration converges to a well-defined limit distribution if there exists only a single fixed point. If this was not the case, the huge number of the iteration variables would cause some fields approaching a solution while the rest another one. Such a simple organization of the solution space is called replica symmetry, and thus the assumption that this iteration converges corresponds to the assumption that the replica symmetry is not broken. Given this, the probability distribution $P_d(h)$ of the fields of vertices of degree d is given by the self-consistency equation

$$P_d(h) = \int_{-\infty}^{\infty} \prod_{l=1}^{d} (dh_l \sum_{d'=0}^{\infty} p(d'|d) P_{d'}(h_l)) \delta(h - \mu - \sum_{l=1}^{d} u(h_l)) \,, \tag{12}$$

in which averaging over the ensemble has been performed.

Let us now turn back to the original problem of vertex covers. The lattice gas of Eq. (5) corresponds to a vertex cover with the choice

$$e^{w(x_i,x_j)} = 1 - x_i x_j \,. \tag{13}$$

In this setting, a particle (a site with $x_i = 1$) corresponds to an uncovered site whereas and a site without a particle to a covered one. Eq. (13) describes a vertex cover since the right-hand side is zero if and only if $x_i = x_j = 1$, i.e. when two adjacent vertices are uncovered. In this case $w(x_i, x_j) = -\infty$ and such configurations are thus not counted in the partition function since $e^{-\beta H} = 0$.

Minimum vertex covers are obtained when the particle number is at its maximum. Therefore, one takes the limit $\mu \to \infty$ with $z = h/\mu$ fixed, substitutes Eq. (13) to Eq. (11), and arrives at

$$P_d(z) = \int_{-\infty}^{\infty} \prod_{l=1}^{d} (dz_l \sum_{d'=0}^{\infty} p(d'|d) P_{d'}(z_l)) \delta(h - \mu - \sum_{l=1}^{d} \max(0, z_l)) \,. \tag{14}$$

By a clever Ansatz, Vazquez and Weigt have been able to solve this equation (see [2] for details). In short, the relative size of the minimum vertex cover χ_c reads

$$\chi_c = 1 - \sum_{d=0}^{\infty} p_d (1 - \pi_{d-1})^{d-1} (1 + \frac{d-2}{2} \pi_{d-1}) \,, \tag{15}$$

where the auxiliary variables π_d obey the self-consistency equation

$$\pi_d = \sum_{d_l=0}^{\infty} p(d_l|d)(1 - \pi_{d_l}^{d_l}) \,. \tag{16}$$

In this solution, the physical interpretation of π_d is that π_d is the probability that an edge arriving at a vertex of degree $d+1$ carries a constraint, i.e. that it is not covered by the neighbouring vertex.

Eq. (16) is a similar fixed point iteration to Eq. (10), except that the former deals with degree classes (subsets of vertices with the given degree d) instead of individual vertices. Nevertheless, convergence of Eq. (16) corresponds to replica symmetry.

IV. NUMERICS

Having arrived at Eqs. (15) and (16), we are ready to set up a protocol of how to compute numerically the minimum vertex-cover size for an ensemble of graphs with given degree distribution and correlations (Eq. (3)). The procedure is as follows. First, iterate Eq. (16). If the iteration converges, substitute the obtained π_d's to Eq. (15) to obtain the relative size of the minimum vertex cover. If the iteration fails to converge, the calculation in the previous section is unable to produce meaningful results, and the conclusion is that the replica symmetry is broken.

To compare the results with numerical experiments on actual graphs, Vazquez and Weigt generate correlated random graphs with a power-law degree distribution $p_d \propto d^{-\gamma}$ using the algorithm outlined in Sec. II. An approximation of the minimum vertex cover is then created using a generalization of the leaf-removal algorithm [8]. In it, at each step the vertex with minimum current degree is chosen, all its neighbours are covered. Then, the considered vertices and all edges emanating from them are removed and the procedure is repeated. If the algorithm produces a vertex cover without ever having to choose an edge with degree $d \geq 2$, i.e. there is always a leaf to be removed with its only neighbour covered, it has found a minimum vertex cover. Otherwise, when choosing a vertex with degree d an error of at most $d-1$ can be caused to the size of the minimum vertex cover. Therefore an upper bound for the error of the algorithm is given by

$$E(\{d_1, \cdots, d_k\}) = \sum_{i=1}^{k} (d_k - 1)(1 - \delta_{d_k,0}) \,, \tag{17}$$

where k is the number of vertices chosen in the course of the computation, and d_k the the degree of the kth chosen vertex at the time of its selection.

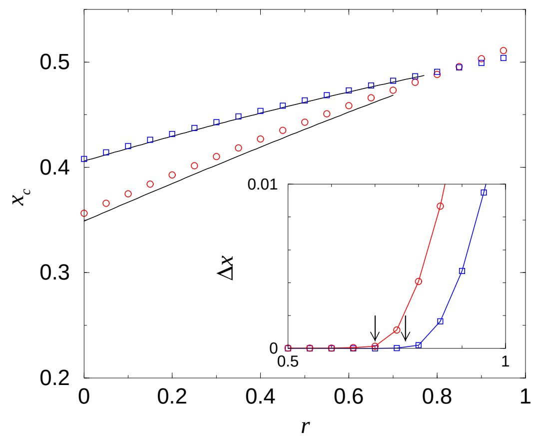

FIG. 2: Main figure: Relative minimum vertex-cover size for a single network with a power-law degree distribution and degree–degree correlations given by Eq. (4). The black solid lines give the analytical solution for $\gamma = 2.5$ (upper curve) and $\gamma = 3.0$ (lower curve). The curves stop at the point where the iteration of Eq. (16) stops converging. The symbols are the numerical estimates for $\gamma = 2.5$ (red circles) and $\gamma = 3.0$ (blue squares). The network size is $N = 10^6$. Inset: The upper bound for the error (Eq. (17)). The figure is from Ref. [2].

The analytical solution is compared to the numerical experiments with the leaf-removal algorithm in Fig. 2. Power-law degree distribution with exponent $\gamma = 2.5$ and $\gamma = 3.0$ has been used together with the correlations defined by Eq. (4). The results show that when the correlation strength parameter r is increased, the iteration of Eq. (16) stops converging at some point. Up to this point, the analytical solution is in quite good (although not perfect – see Sec. V) agreement with the numerical experiments. Above that, these two cannot be compared since the analytical treatment fails to give a solution.

In the inset of Fig. 2, the upper bound for the error of the leaf-removal algorithm (Eq. (17)) is plotted against r in the same two example cases. The error starts deviating from zero at the same point where Eq. (16) ceases to converge. The extensive cumulation of error of the leaf-removal algorithm has been associated with replica-symmetry breaking. Thus, the fact that the failure to converge and the onset of the nonzero error occur at the same point support the conclusion that the point is associated with replica-symmetry breaking. There is, however, a suspicious discrepancy between the main figure and the inset. Namely, the difference between the analytical result and the numerical experiments in the main figure is clearly at least an order of magnitude greater than the algorithmic error in the inset. The authors of Ref. [2] state that this is due to “finite-size corrections resulting mainly from a degree cutoff” . For another possible explanation, see the discussion in the next section.

V. DISCUSSION

The results have some consequences regarding applications. First, the benchmark networks used have properties in common with real-world networks. Fat tails of the degree distribution have been widely observed (see, for instance, [4]), and the power-law form serves as a good approximation for them. A reasonable example is the topology of the Internet at the autonomous systems level, where power-law degree distributions have been measured with $\gamma \approx 2.2$ [9]. Real-world graphs are often also correlated. Assortative correlations, i.e. the tendency of the vertices to be connected to vertices with a similar degree, are usually associated with social networks whereas disassortative correlations are associated with technological networks and those related to traffic [10].

Second, one application of minimum vertex covers is network traffic monitoring. Here, one wants to deploy observation points on a network in which vertices are routers and edges the physical links connecting the routers, such that each edge can be directly observed [5]. The least costly setup of the observation points is therefore the minimum vertex cover, and a step in such studies is to compute or estimate it. The results of Vázquez and Weigt point out that computing the minimum vertex cover should be easy for such graphs with the generalized leaf-removal algorithm since the correlations in such nets are typically negative (r is negative).

The statement in the previous paragraph is not, however, completely true. In technological nets, the correlations are often disassortative [10]. In the current setting, this would translate to negative values of r (since at $r = 0$ there are no correlations). However, this regime cannot be obtained directly by the means utilised by Vázquez and Weigt, since in the network generation procedure, r has an interpretation as a probability and thus it has to be non-negative. Widening the approach to deal with negative r could be an interesting starting point for further studies.

One has to be also careful with generating correlated graphs especially when fat-tailed degree distributions are involved. Eq. (4) certainly describes the creation procedure of the networks, but it is not immediately clear that it consequently describes the correlations in the resulting network. To see this, consider creating a power-law random graph where correlations are absent. At first glance, one would just set $e_{dd'} = q_d q_{d'}$ instead of Eq. (4) and proceed as above. However, Catanzaro *et al.* have pointed out that this, in fact, results in correlated graphs due to topological constraints [11]. In other words, the fact that all stubs of highly connected nodes have to be connected somewhere creates limitations in the overall connection structure, which, in turn, produces degree–degree correlations. This can be overcome by forcing a suitable upper cutoff for the degree [11]. Unless this phenomenon has been taken into account by the authors of Ref. [2], it provides another possible explanation for the discrepancy between Fig. 2 and its inset.

---

[1] A. K. Hartmann and M. Weigt, *Phase Transitions in Combinatorial Optimization Problems: Basics, Algorithms and Statistical Mechanics*, Wiley VCH (2005).
[2] A. Vázquez and M. Weigt, Phys. Rev. E **67**, 027101 (2003).
[3] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[4] S. N. Dorogovtsev and J. F. F. Mendes, Adv. in Phys. **51**, 1079 (2002).
[5] Y. Breitbart, C.-Y. Chan, M. Garofalakis, R. Rastogi, and A. Silverschatz, Proceedings of IEEE INFOCOM, Alaska, 2001.
[6] B. Bollobás, *Random graphs*, Academic Press, New York (2001).
[7] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[8] M. Weigt, Eur. Phys. J. B **28**, 369 (2002).
[9] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, IEEE/ACM Transactions on Networking **4**, 514 (2003).
[10] M. E. J. Newman, Phys. Rev. E. **67**, 026126 (2003).
[11] M. Catanzaro, M. Boguña, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).