GRAPH INVARIANT FROM IDEAS IN QUANTUM FIELD THEORY

AN HUANG, SHING-TUNG YAU

Abstract. We propose a polynomial time computable graph invariant, which comes out directly from a modified version of discrete Green's function on a graph. It is our hope that it can help to construct fast algorithms for the graph isomorphism testing. We explain the physics behind this discrete Green's function, which is a very basic idea applied to graphs.

1. INTRODUCTION

The graph isomorphism problem is a long standing problem that is of both theoretical and practical importance. For complexity considerations, the problem is clearly in NP, but it is neither known to be in P nor NP-complete. In practice, it has many applications, in e.g. chemistry, image processing, etc. Due to its importance, much effort has been put into the research of this problem, and a lot have been learned: e.g. for random graphs, or graphs with various special properties, or graphs coming from practice in various specific ways, there are various known fast algorithms. Many graph invariants have been studied, and put into work. However, as stated in [1], it is clear that there lacks a uniform and deeper understanding of this problem, thus causing many issues in an unclear stage. The main motivation of this paper is to try to initiate a new perspective to the study of graphs, and in particular to the graph isomorphism problem, from ideas familiar in physics, with the hope of eventually providing a deeper understanding.

More concretely, in section 2, we propose a particular polynomial time computable graph invariant. In section 3, we study one of the simplest quantum field theories defined on a graph[1], namely a real free scalar field theory, with a varying mass parameter. Its two point correlation function gives us a version of the discrete Green's function, which determines the graph up to isomorphism. This function showed up in [7] for different purposes. The graph invariant we propose, arises directly from this discrete Green's function. In section 4, we study some basic properties of this graph invariant. In particular, it turns out that if two cospectral graphs share this invariant, then infinitely many identities have to be satisfied, which should strongly constrain the possibilities. Thus it looks that very basic physics idea may produce interesting graph invariants.

The invariant we propose belongs to the framework of spectral graph theory: it is constructed using eigenspaces of the Laplacian matrix. As stated in e.g. [2][4], there is a hope to discover good complete invariants from this approach. Our idea is related with the idea of graph angles that is surveyed in e.g. [4][5]. [3] is also of relevance to our idea, where the authors use the eigenspace to constrain the action of the automorphism group of the graph, on the coset space of the eigenspace.

---

[1]Here we use the notion quantum field theory in a sense similar to lattice gauge theory: we apply some of its very basic ideas, in a situation where there are only finitely many degrees of freedom.

2. A GRAPH INVARIANT

Let G be a graph with $|G| = n$ vertexes, choose an arbitrary labeling of the vertexes by $x_1, x_2, ..., x_n$, and let M denote its $n \times n$ (combinatorial) Laplacian matrix under this basis: for $i \neq j$, the i, j-th entry is equal to -1 if there is an edge between x_i and x_j, and is equal to 0 otherwise. The diagonal entries are the degrees of the vertexes, so that the sum of any column of M is equal to 0. From definition, M is symmetric. M represents the combinatorial Laplacian operator under the dual basis.

Remark 2.1. The following method is linear algebra, that can also work for suitable variations of the Laplacian matrix, e.g. the normalized Laplacian. Furthermore, the discussion can actually be applied to more general situations, such as multigraphs.

Suppose we have another graph G_1 with n vertexes, and upon a choice of an arbitrary labeling of the vertexes, we get another Laplacian matrix M_1. The problem of whether G and G_1 are isomorphic graphs, amounts to the linear algebra question of whether there exists a permutation matrix P, such that $P^t M P = M_1$. (Note that $P^t = P^{-1}$) In spectral graph theory, people study the real spectrum of M, as an invariant of the graph under isomorphisms, however, the spectrum itself is not sufficient for the graph isomorphism problem: two graphs can have the same real spectrum but fail to be isomorphic, and these are called cospectral graphs. On the other hand, the eigenfunctions contain much more information than just the eigenvalues. The apparent question of dealing with the eigenfunctions or eigenspaces is that, they are not preserved under graph isomorphisms but instead, the eigenspaces also transform by permutations. So one needs to find suitable invariants associated with the eigenfunctions, in order to use them appropriately in the graph isomorphism problem. To do this, we first normalize the eigenfunctions in the standard way: choose an orthonormal basis for each (real) eigenspace, in the sense that they are all orthogonal to each other, and the L^2 norm of each is 1: namely, let λ_k, $k = 1, 2, ..., m$ denote the set of different eigenvalues of M by increasing order. For each k, let the column vectors $\phi_k^1, ..., \phi_k^{l_k}$ denote an orthonormal basis of the corresponding eigenspace E_k.

Denote $t_k(x, y) = \sum_{i=1}^{l_k} \phi_k^i(x)\phi_k^i(y)$, and $T(x, y) = \langle t_1(x, y), ..., t_m(x, y) \rangle$. It is obvious that the vector function $T(x, y)$ does not depend on the choice of the orthonormal basis, and it can be constructed directly from the graph Laplacian independent of the choice of a labeling of vertexes, therefore it is an intrinsically defined function on $G \times G$. The set of $1 \times m$ vectors $T(x, y)$ counting multiplicity, where x, y range among all pairs of vertexes of G, which we name by ST, is therefore an invariant of the graph, which is clearly polynomial time computable, and furthermore the elements of this set can be ordered in order for comparisons.

3. PHYSICS INTERPRETATION

Consider an Euclidean real scalar field theory on the graph G: the space of fields is then the space of all real valued functions on vertexes of G, which is an n dimensional real vector space. We write the free field Lagrangian with a mass parameter $u = m^2$ in direct analogy with the familiar Lagrangian in the continuous situation:

$$(3.1) \qquad \mathfrak{L} = \sum_{e \in E} (\nabla_e \phi)^2 + u\phi^2$$

where ∇_e is the graph gradient with respect to a directed edge e, $(\nabla_e \phi)^2$ is independent of the choice of the orientation of e, and E is the set of edges of G. One can consult e.g. [8] for these notations. We have the usual Green's formula

$$(3.2) \qquad \int_G (\nabla_e \phi)^2 dx = \int_G \phi \Delta\phi dx$$

where Δ is the Laplacian.

As the same with usual QFT on a manifold, we consider two point correlation functions defined as

$$(3.3) \qquad \langle\phi(x)\phi(y)\rangle = \frac{\int \phi(x)\phi(y)e^{-\int \phi(\Delta+u)\phi dx}D\phi}{e^{-\int \phi(\Delta+u)\phi dx}D\phi}$$

We allow x and y to be equal, as there will be no short distance problems in our situation. This is a finite dimensional path integral of the type that is often used as toy model to introduce the Feynman rules in physics textbooks, and it is free of divergences. However, in our simple situation here, this is our path integral. We know very well how to evaluate this by undergraduate calculus with familiar result: the denominator equals the determinant of the Laplacian to the power $-\frac{1}{2}$, which cancels with a factor coming from the numerator. Up to a nonzero constant scalar, what is left is a sum over different eigenvalues of the form

$$(3.4) \qquad \sum_{k=1}^{m} \frac{t_k(x,y)}{\lambda_k + u}$$

Which may be viewed as a discrete version of the Fourier transform of the D'Alembert propagator, the familiar result in usual QFT. The individual $t_k(x,y)$ for each eigenvalue may be recovered as residues near different poles of the two point correlation function, as we vary the parameter u.

It is straightforward to check that the function $\langle\phi(x)\phi(y)\rangle$ satisfies a discrete version of the quantum equation of motion

$$(3.5) \qquad L_x \langle\phi(x)\phi(y)\rangle = \delta_{x,y}$$

where L_x is the Laplacian operating on coordinate x, and the delta function $\delta_{x,y}$ on a graph is given by

$$\delta_{x,y} = \begin{cases} 0, & x \neq y, \\ 1, & x = y. \end{cases}$$

Therefore, we call the two point correlation function by the discrete Green's function, and upon a choice of labeling of vertexes as we have done, (3.5) becomes the statement that, $\langle\phi(x)\phi(y)\rangle$ as a matrix, is the inverse of $M + uI$. So obviously, it determines the graph up to isomorphism.

Remark 3.1. The two point correlation function determines the graph up to isomorphism, thus it also determines the QFT on the graph, and therefore all of its correlation functions. This can be viewed as a baby version of Wick's theorem in the graph case.

Furthermore, one can then study various operations on graphs, and try to see how the two point correlation function changes accordingly. This is interesting because, theoretically, it is almost always important to understand how invariants change under important operations. On the other hand, the two point correlation function as a complete invariant may provide a measure on when two given graphs are considered "almost isomorphic", which may be useful in practice– e.g. if we have a large data presented as a big graph, one should expect that the data given may contain a little marginal error, and so being able to make sense of and detect "almost isomorphic" graphs looks to be a practically important problem. It looks interesting to investigate whether the two point function or something similar can give practically useful definitions of almost isomorphic graphs.

For example, suppose we delete an edge (adding an edge will be just the opposite, of course) between two vertexes x_1 and x_2, and get a new graph we call G_2. Let us try to write down the two point correlation function for G_2, in terms of data of G and the two vertexes x_1 and x_2. From the form of (3.3), we know that this operation may only possibly affect the

term $e^{\int \phi\Delta\phi dx}$. For this term, at any vertex other than x_1 and x_2, the action of the Laplacian is unaffected by definition. At x_1, the integral $\int \phi\Delta\phi dx$ changes by $\phi(x_1)(\phi(x_1) - \phi(x_2))$, and at x_2, the integral changes by $\phi(x_2)(\phi(x_2)-\phi(x_1))$. Therefore, the two point correlation function for G_2 can be expressed as

$$(3.6) \qquad \langle\phi(x)\phi(y)\rangle_{G_2} = \frac{\int \phi(x)\phi(y)e^{\int \phi\Delta\phi dx}e^{(\phi(x_1)-\phi(x_2))^2}D\phi}{e^{\int \phi\Delta\phi dx}e^{(\phi(x_1)-\phi(x_2))^2}D\phi}$$

Again, the above can be explicitly calculated by Gaussian integrals, and one may then compare it with the two point correlation function of G, and analyze the difference in various situations. One elementary observation is that, roughly speaking, difference of values of eigenfunctions at vertexes x_1 and x_2 contribute to the difference of two point correlation functions, and furthermore, the two point correlation function is more sensitive to the difference at smaller eigenvalues. This is consistent with the physics picture: smaller eigenvalues correspond to lower energy modes, and if the low energy modes for two graphs are close, then we have a sense that these two graphs are close to each other.

Remark 3.2. The individual functions $t_k(x, y)$ will change in a more complicated manner, and probably one should not expect a particularly nice formula for the change of $t_k(x, y)$ similar to (3.6), because e.g. even the number of eigenvalues and the dimension of eigenspaces may jump, and there may be complications from cross terms. The combination $\langle\phi(x)\phi(y)\rangle$ takes into account all of these and the change of it can be presented by the simple formula above.

Note that the operation of deleting or adding an edge can turn any graph with a given finite number of vertexes to any other graph with the same number of vertexes, it is our hope that one can then try to estimate further how two point correlation functions change, and develop useful quantitative concepts for "almost isomorphism" from this. Of course, there may well be other reasonable concepts of "almost isomorphism", but our guideline here is that a concept coming naturally from physics may be worth of study. More generally, it is conceivable that one may also use the two point correlation function to study other operations or problems on graphs.

Remark 3.3. It looks quite possible that one may study more elaborated quantum field theories on a general graph, especially with the help of various topological and geometrical concepts for graphs that are developed for graphs recently.

4. A FIRST STUDY OF THE INVARIANT

As the two point function matrix is the inverse of $M + uI$, by the adjugate matrix formula of an inverse matrix, we have

$$(4.1) \qquad \langle\phi(x)\phi(y)\rangle = \frac{(-1)^{x+y}A_{y,x}}{det(M + uI)}$$

where $A_{y,x}$ is the y, x-th cofactor of $M+uI$, which is a polynomial in u of integral coefficients of degree less than n. Since our discrete Green's function can be written as an integral of the heat kernel which is positive, one expects $\langle\phi(x)\phi(y)\rangle$ to be positive. In fact, one has the following stronger fact

Lemma 4.1. *All coefficients of the polynomial* $(-1)^{x+y}A_{y,x}$ *are positive.*

Proof. This is a simple verification by induction. $\square$

We consider the graph invariant given by the set of values (actually a set of functions of u) of the two point correlation function, counting multiplicities, together with the multiplicity

information of the real spectrum. We have

$$(4.2) \qquad \langle \phi(x)\phi(y)\rangle = \sum_{k=1}^{m}\sum_{i=1}^{l_k}\frac{\phi_k^i(x)\phi_k^i(y)}{\lambda_k + u}$$

and by basic linear algebra, more generally,

$$(4.3) \qquad (M + uI)^{\alpha} = \sum_{k=1}^{m}\sum_{i=1}^{l_k}(\phi_k^i(x)\phi_k^i(y))(\lambda_k + u)^{\alpha}$$

for any $\alpha \in \mathbb{R}$. Note that one can take such arbitrary powers of a positive semi-definite matrix.

Therefore, if for two cospectral graphs G and G_1, the invariant we are considering are the same, this means that there exists a permutation T of n^2 elements acting linearly on $n \times n$ matrices by permuting the corresponding elements, such that

$$(4.4) \qquad T\langle\phi(x)\phi(y)\rangle_G = \langle\phi(x)\phi(y)\rangle_{G_1}$$

By the above equation combined with taking residues of (4.2), we have for all k,

$$(4.5) \qquad T\left\langle\sum_{i=1}^{l_k}\phi_k^i(x)\phi_k^i(y)\right\rangle_G = \left\langle\sum_{i=1}^{l_k}\phi_k^i(x)\phi_k^i(y)\right\rangle_{G_1}$$

Therefore by (4.3), we have

$$(4.6) \qquad T(M + uI)^{\alpha} = (M_1 + uI)^{\alpha}$$

for all $\alpha \in \mathbb{R}$.

(4.6) looks to be a rather strong condition on the Laplacian matrix, as it gives (uncountably) infinitely many identities of functions of u. e.g. Taking $\alpha = 0$, one derives that T preserves the diagonal. Taking α to be positive integers, and $u = 0$, one gets infinitely many identities with more or less clear combinatorial meaning.

One the other hand, as it is clear from the above derivation, the set of $1 \times m$ vectors $T(x,y)$, which we denoted by ST, together with the real spectrum, as an invariant of the graph, is equivalent to the above set of values of two point correlation functions.

Remark 4.1. It may be expensive to compute the cofactors $A_{y,x}$ as a polynomial in u. However, if one prefers to work with positive integers, one can instead take n different integral values of u, and compute the corresponding n positive integral values of $A_{y,x}$, which determine $A_{y,x}$ uniquely as it is a polynomial of degree less than n. This can be done in polynomial time, however, the price is that very possibly, in this way one has to deal with very big integers.

Motivated by (4.6), we have the following obvious question:

Question 4.1. Is the invariant we discussed, namely ST together with the real spectrum, a complete invariant of graphs?

If the answer is yes, as a consequence then the graph isomorphism problem is in P. If the answer is no, it will be of interest to find a counterexample. In any case, since this gives an easily computable and comparable graph invariant which also looks rather strong, it should be of interest to investigate more, and also to optimize practical algorithms based on it.

As it is mentioned in the beginning of the paper, this method can also work for some other variants of the Laplacian matrix. As an example, we can consider the normalized Laplacian and calculate the same invariant. In this situation, the cofactors are no longer polynomials with integral coefficients, but that is not essential. The key facts such as equation (4.6) still

hold, and therefore one again gets a graph invariant that looks rather strong. On the other hand, there are well-known examples of large sets of cospectral graphs w.r.t the normalized Laplacian that one can test on.

References

[1] S. Fortin, *The Graph Isomorphism Problem*, thesis, University of Alberta, 1996
[2] D. Cvetkovic, M. Doob, H. Sachs, A. Torgasev, *Recent Results in the Theory of Graph Spectra*, Annals of Discrete mathematics (36), 1988
[3] L. Babai, D. Grigoryev, M. Mount, *Isomorphism of graphs with bounded eigenvalue multiplicity*, Proceedings of the 14th Annual ACM Symposium on Theory of Computing, pp. 310-324, 1982
[4] D. Cvetkovic, P. Rowlinson, S. Simic, *Eigenspaces of Graphs*, Encyclopedia of Math. and its applications (66), 1997
[5] edited by L. Beineke, R. Wilson, *Topics in Algebraic Graph Theory*, Encyclopedia of Math. and its applications (102), 2004
[6] J. Conway, *The Sensual (Quadratic) Form*, Carus Mathematical Monographs, No. 26, 1997
[7] F. Chung, S.-T. Yau, *Discrete Green's functions*, Journal of combinatorial theory (A), 91, (2000), 191-214
[8] F. Chung, A. Grigoryan, S.-T. Yau, *Higher eigenvalues and isoperimetric inequalities on Riemannian manifolds and graphs*, Comm. Anal. Geom. 8 (2000) no.5, 969-1026

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY, CAMBRIDGE, MA 02138.