論文の内容の要旨

論文題目：Computational and chemical barriers for counting-dependent characterization of biomolecular networks
（計数に基づく生体分子ネットワークの特性評価における計算論および化学論的困難）

論文提出者氏名：　バリッシュ　ロバート　ダニエル

In this dissertation we concern ourselves with the “hardness”, in the sense of both computational complexity theory as well as biochemistry or chemistry formal, of a few difficult counting problems that have direct relevance to the characterization of biochemical systems, and in particular, metabolic networks. Along the way we also prove a number of theories that may be of independent interest for computer scientists and graph theorists.

In Chapter 2 we consider a number of complexity theoretic question broadly concerning the hardness of counting and approximating directed as well as undirected Hamiltonian cycles, Hamiltonian paths, and general cycles in graphs with a range of severe vertex degree, partiteness, vertex connectivity, and planarity constraints. Here, we are motivated by the direct correspondence of these complexity results to the problem of counting and approximately counting simple cycles in “reaction-centric graphs” of chemical or biochemical networks, where vertices correspond to “reaction centers” (e.g. enzyme proteins) and where directed (resp. undirected) edges correspond to irreversible (resp. reversible) flows of metabolites between these reaction centers.

In Chapter 3 we extend the results of the previous chapter to show that

counting undirected simple cycles on ($k \geq 3$)-regular (bipartite or planar) graphs, which in the context of the previous chapter can be interpreted as substrate cycles embedded on a ($k \geq 3$)-regular reaction-centric graph at the limit where all reactions are reversible, remains #P -complete. We also prove that counting "circuits" (i.e. walks on graphs where edges are traversed at most once and vertices may be traversed an arbitrary number of times) on ($k = \{3, 4, 5\}$)-regular (bipartite or planar) graphs is #P -complete, and show that this result can be extended to show that counting circuits on ($k \geq 3$)-regular (bipartite or planar) graphs is #P -complete if a polynomial-time formula for circuits on the complete bipartite graph and a complete bipartite graph with one missing edge can be found. However, as an important caveat here is that, as a consequence of our method of proof, we are unable to say anything meaningful regarding the existence or non-existence of approximation algorithms for counting cycles and circuits outside of the $k = 3$ case. We also leave it as an open question to the biological community if there is value in counting or enumerating circuits on ($k \geq 3$)-regular (bipartite or planar) graphs to consider metabolic networks.

An additional point of biochemical motivation for this chapter comes from the field of protein folding. Here we abstract the polymer under consideration as a self-avoiding embedding of a cycle or path of vertices into a graph, and simply require such that each folding of the polymer corresponds to a Hamiltonian cycle or path on the graph. Therefore, our results in this chapter are relevant to the calculation of configurational entropy of protein folding models. We are potentially able to say something interesting in the sense that calculating configurational entropy fails to become easier in the limit of high vertex degrees, which typically correspond to coordination numbers in discrete models of protein folding.

In Chapter 4 we prove a number of results concerning the computational complexity of counting and approximately counting Hamiltonian cycles and paths on highly restricted variants of cubic graphs. Perhaps most notably, we

prove that counting either Hamiltonian paths or cycles on cubic 3-connected bipartite planar graphs is #P -complete and polynomial-time inapproximable unless NP = RP. This is interesting in the sense that the Hamiltonicity of this family of graphs is a major open problem known as Barnette's conjecture, and because it is known that deciding the existence of a Hamiltonian cycle on this family of graphs is NP -complete if and only if Barnette's conjecture is false. Therefore, this is the first case we know of where a #P -completeness result has been achieved corresponding to a decision problem of unknown complexity.

The immediate "biochemically relevant" application of this chapter is that it establishes the strongest case of the Chapter 2 Theorem 2 #P -completeness and inapproxibility result for counting substrate cycles or elementary fundamental modes, at the limit where all reactions are reversible, on cubic 3-connected bipartite planar graphs. In the manner of our "biochemically justification" for Chapter 3, the results in this chapter can also be understood to imply a hardness result for computing the configurational entropy of discrete state approximations of self-avoiding cyclical or linear polymers at the limit where the polymer is in a compact phase. We have that computing the configurational entropy of the polymer under such a model is #P -complete even on cubic 3-connected bipartite planar graphs and polynomial-time inapproximable on this family of graphs unless NP = RP.

In Chapter 5 we attempt to strengthen the results of Chapter 2 and Chapter 4 by showing that the hardness of determining only the least significant bit (i.e. the parity) of the number of Hamiltonian cycles and simple cycles more generally on cubic weakly-3-connected bipartite planar digraphs and subcubic 2-connected bipartite planar undirected graphs, are complete for the class $\oplus$P, and are thus among the hardest known parity counting problems. While we leave it as an open question to the biochemical community as to whether these results have any direct physical significance,

we argue that they provide a strong clue as to the difficulty of efficiently extracting information about the number of cycles in even very topologically restricted classes of graphs.

Finally, in Chapter 6 we attempt a different sort of "worst case" hardness analysis, and consider the physical barriers to counting ultra-dilute RNA species from the perspective of fundamental nucleic acid chemistry. Here, we focus on a series of ultra-high sensitivity assays for RNA and protein species in the cell or tissue sections, and in particular the photo-DEAN method of Yokomori et. al. for counting individual RNA species, and look carefully at the barriers at the limit of low RNA molecular concentrations. Specifically, we look at the difficulty of diffusion-based search in the low target copy number limit; we look at challenges presented by spontaneous decomposition of (deoxy)ribonucleic acid due to depurination or depyrimidination followed by intramolecular cleavage via β-elimination at abasic sites; we look at cytosine deamination via hydrolytic deamination; we look at spontaneous decomposition of RNA via transesterification; finally, we consider cyclobutane pyrimidine dimer formation and other undesired photochemistry as a result of UV irradiation for the photo-DEAN method.