

De Bruijn graph

From Wikipedia, the free encyclopedia

In graph theory, an n -dimensional **De Bruijn graph** of m symbols is a directed graph representing overlaps between sequences of symbols. It has m^n vertices, consisting of all possible length- n sequences of the given symbols; the same symbol may appear multiple times in a sequence. If we have the set of m symbols $S := \{s_1, \dots, s_m\}$ then the set of vertices is:

$$V = S^n = \{(s_1, \dots, s_1, s_1), (s_1, \dots, s_1, s_2), \dots, (s_1, \dots, s_1, s_m), (s_1, \dots, s_2, s_1), \dots, (s_m, \dots, s_m, s_m)\}.$$

If one of the vertices can be expressed by shifting all symbols by one place to the left and adding a new symbol at the end of another vertex, then the latter has a directed edge to the former vertex. Thus the set of (directed) edges is

$$E = \{((v_1, v_2, \dots, v_n), (w_1, w_2, \dots, w_n)) : v_2 = w_1, v_3 = w_2, \dots, v_n = w_{n-1}\}.$$

Although De Bruijn graphs are named after Nicolaas Govert de Bruijn, they were discovered independently by both De Bruijn^[1] and I. J. Good^[2]. Much earlier, Flye Sainte-Marie^[3] implicitly used their properties.

Some popular grid network topologies are De Bruijn graphs. The Koorde protocol uses a De Bruijn graph.

In bioinformatics De Bruijn graphs are used for de novo assembly of short read sequences into a genome^[4] ^[5] ^[6].

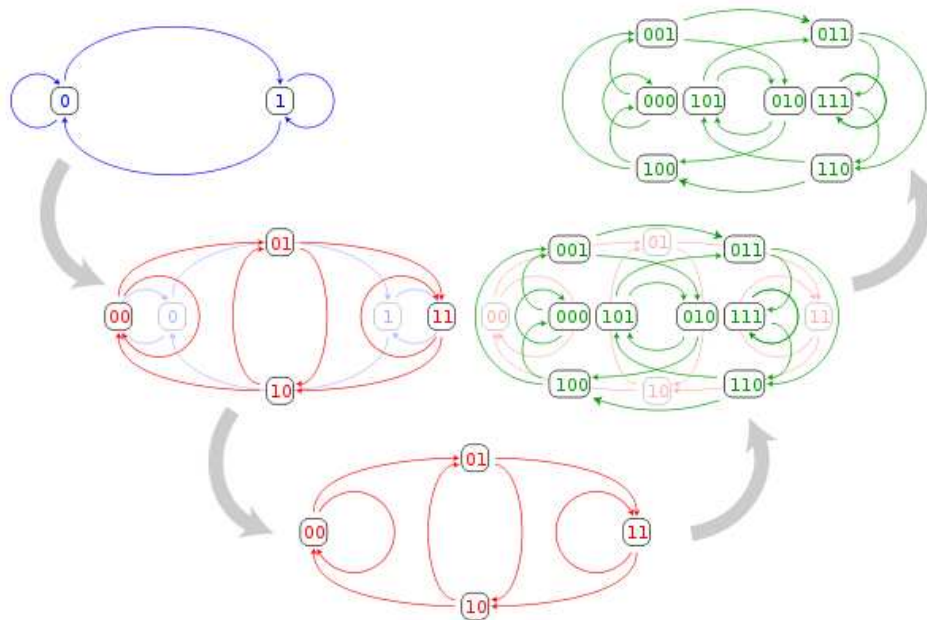
Contents

- 1 Properties
- 2 Dynamical systems
- 3 Uses
- 4 See also
- 5 References

Properties

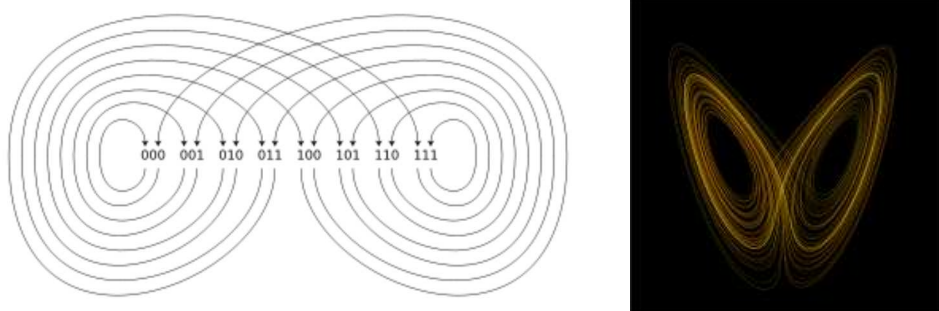
- If $n = 1$ then the condition for any two vertices forming an edge holds vacuously, and hence all the vertices are connected forming a total of m^2 edges.
- Each vertex has exactly m incoming and m outgoing edges.
- Each n -dimensional De Bruijn graph is the line digraph of the $(n - 1)$ -dimensional De Bruijn graph with the same set of symbols^[7].
- Each De Bruijn graph is Eulerian and Hamiltonian. The Euler cycles and Hamiltonian cycles of these graphs (equivalent to each other via the line graph construction) are De Bruijn sequences.

The line graph construction of the three smallest binary De Bruijn graphs is depicted below. As can be seen in the illustration, each vertex of the n -dimensional De Bruijn graph corresponds to an edge of the $(n - 1)$ -dimensional De Bruijn graph, and each edge in the n -dimensional De Bruijn graph corresponds to a two-edge path in the $(n - 1)$ -dimensional De Bruijn graph.



Dynamical systems

Binary De Bruijn graphs can be drawn (below, left) in such a way that they resemble objects from the theory of dynamical systems, such as the Lorenz attractor (below, right):



This analogy can be made rigorous: the n -dimensional m -symbol De Bruijn graph is a model of the Bernoulli map

$$x \mapsto mx \bmod 1$$

The Bernoulli map (also called the $2x \bmod 1$ map for $m = 2$) is an ergodic dynamical system, which can be understood to be a single shift of a m -adic number^[8]. The trajectories of this dynamical system correspond to walks in the De Bruijn graph, where the correspondence is given by mapping each real x in the interval $[0,1)$ to the vertex corresponding to the first n digits in the base- m representation of x . Equivalently, walks in the De Bruijn graph correspond to trajectories in a one-sided subshift of finite type.

Uses

- Some grid network topologies are De Bruijn graphs.
- The distributed hash table protocol Koorde uses a De Bruijn graph.
- In bioinformatics De Bruijn graphs are used for de novo assembly of (short) read sequences into a genome^[4] ^[5]

See also

- De Bruijn sequence
- Kautz graph
- Free monoid
- Semiautomata

References

- [^] de Bruijn, N. G. (1946). "A Combinatorial Problem". *Koninklijke Nederlandse Akademie v. Wetenschappen* **49**: 758–764.
- [^] Good, I. J. (1946). "Normal recurring decimals". *Journal of the London Mathematical Society* **21** (3): 167–169. doi:10.1112/jlms/s1-21.3.167 (http://dx.doi.org/10.1112%2Fjlms%2Fs1-21.3.167) .

3. ^ Flye Sainte-Marie, C. (1894). "Question 48". *L'Intermédiaire Math.* **1**: 107–110.
4. ^ ^a ^b Pevzner, Pavel A.; Tang, Haixu; Waterman, Michael S. (2001). "An Eulerian path approach to DNA fragment assembly". *PNAS* **98** (17): 9748–9753. Bibcode 2001PNAS...98.9748P (http://adsabs.harvard.edu/abs/2001PNAS...98.9748P) . doi:10.1073/pnas.171285098 (http://dx.doi.org/10.1073%2Fpnas.171285098) . PMC 55524 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=55524) . PMID 11504945 (http://www.ncbi.nlm.nih.gov/pubmed/11504945) .
5. ^ ^a ^b Pevzner, Pavel A.; Tang, Haixu (2001). "Fragment Assembly with Double-Barreled Data". *Bioinformatics/ISMB* **1**: 1–9.
6. ^ Zerbino, Daniel R.; Birney, Ewan (2008). "Velvet: algorithms for de novo short read assembly using de Bruijn graphs". *Genome Research* **18** (5): 821–829. doi:10.1101/gr.074492.107 (http://dx.doi.org/10.1101%2Fgr.074492.107) . PMC 2336801 (http://www.pubmedcentral.gov/articlerender.fcgi?tool=pmcentrez&artid=2336801) . PMID 18349386 (http://www.ncbi.nlm.nih.gov/pubmed/18349386) .
7. ^ Zhang, Fu Ji; Lin, Guo Ning (1987). "On the de Bruijn-Good graphs". *Acta Math. Sinica* **30** (2): 195–205.
8. ^ Leroux, Philippe (2002), *Coassociative grammar, periodic orbits and quantum random walk over Z*, arXiv:quant-ph/0209100 (http://arxiv.org/abs/quant-ph/0209100) , Bibcode 2002quant.ph..9100P (http://adsabs.harvard.edu/abs/2002quant.ph..9100P)

Retrieved from "http://en.wikipedia.org/wiki/De_Bruijn_graph"

Categories: Dynamical systems | Automata theory | Parametric families of graphs | Directed graphs

- This page was last modified on 5 June 2011 at 00:42.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.

Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.