

Synthesis Digraphs and their Vulnerability. The Risk of Failure of Synthesis Plans

Steven H. Bertz

AT & T Bell Laboratories, Murray Hill, New Jersey 07974, U.S.A.

Synthesis graphs are redefined as *digraphs* (directed graphs) which leads to the consideration of the 'vulnerability' of synthetic plans and its relationship to the risk of failure upon execution of the plan.

As in the usual model,^{1,2} the points of the synthesis graph† are taken to represent chemical compounds‡ (starting materials S, intermediates I, targets T) and the lines are taken to represent reactions converting one compound into another. Before proceeding to the main results, however, it must be pointed out that synthesis plans should be represented by directed

graphs or *digraphs*,³ in which the points are joined by *directed lines*, (*i.e.*, arrows) to indicate the directions in which the reactions proceed.

Then, it is possible to introduce the concept of *vulnerability*, loosely speaking, the effect of failure of projected reactions. To give this concept more precise definition, the number of *vulnerable lines*, $l_v(S, T)$, counts those lines in paths from S to T such that there is no other path (of any length) from the origin of the line to its terminus.§ An especially vulnerable kind of line is one that, if removed, results in a disconnected graph. Such a line is called a *bridge*,³ and the number of bridges in paths from S to T is a more critical measure of vulnerability than $l_v(S, T)$, since the removal of a bridge makes

† Previously called the 'synthetic tree' or 'synthesis tree' by Corey (ref. 1) and termed the 'planning graph' by Hendrickson (ref. 2). Corey's tree was drawn in the 'retrosynthetic' direction and had some arrows indicating '*etc.*', but it did not use directed lines between structures.

‡ A point usually represents a single compound; nevertheless, it can represent a mixture of, *e.g.*, enantiomers, starting materials, or products where this is deemed desirable.

§ In ref. 3b these are termed basic lines.

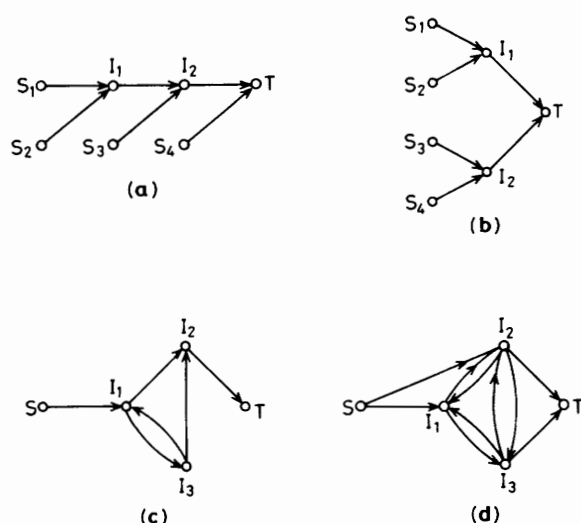


Figure 1. Some typical synthesis plans represented as digraphs.

it impossible to proceed from S to T. In Figures 1(a) and 1(b) respectively, l_v (linear) = l_v (convergent) = 6, *i.e.*, every line is a vulnerable line; in fact, every line is a bridge. In Figure 1(c), $l_v = 4$ (only $I_1 \rightarrow I_2$ and $I_3 \rightarrow I_2$ are not vulnerable lines[¶]) of which two are bridges ($S \rightarrow I_1$ and $I_2 \rightarrow T$). There are neither vulnerable lines nor bridges in Figure 1(d).

A related measure of vulnerability is the minimum number of lines which, if removed, makes it impossible to proceed from S to T (*e.g.*, from starting material to target). By a variation of Menger's theorem,^{||} this equals the maximum number of *line-disjoint paths* from S to T, denoted by $p_{ld}(S, T)$. Line-disjoint paths are those which have no lines in common, although they may share points. Going back to Figures 1(a) and 1(b), $p_{ld}(\text{lin.}) = p_{ld}(\text{conv.}) = 1$, *i.e.*, the failure of one reaction will scuttle either a linear or a convergent synthesis. In Figure 1(c), $p_{ld}(S, T) = 1$ as well, even though the numbers of bridges and vulnerable lines are lower. In Figure 1(d), $p_{ld}(S, T) = 2$, and at least two reactions must fail for the synthesis to fail to reach the target. The simplest way to ensure that $l_v = 0$ and that $p_{ld}(S, T) > 1$ is to convert the synthesis graph into a *multigraph*³ with two or more (directed) lines from one point to another. Then there are back-ups in case reactions fail.

Some of the concepts defined above for lines (reactions) have analogies when points (compounds) are considered. For example, a point that leaves a disconnected graph when it is removed is called a *cutpoint*³ (*cf.* bridge), also called a 'separation vertex' or 'articulation point.'⁴ Cutpoints in a synthesis graph should be noted, and the corresponding compounds should be scrutinized very carefully to make sure that they are energetically feasible. Two paths are *point-disjoint* if they have no points other than their endpoints in common (*cf.* line-disjoint paths). According to Menger's theorem,^{||} the maximum number of point-disjoint paths from S to T, $p_{pd}(S, T)$, equals the minimum number of points that must be removed to separate S and T.

[¶] $I_1 \rightarrow I_2$ can be replaced by the path $I_1 \rightarrow I_3 \rightarrow I_2$; $I_3 \rightarrow I_2$ can be replaced by $I_3 \rightarrow I_1 \rightarrow I_2$.

^{||} See ref. 3a, p. 47 and ref. 3b, p. 247.

All the intermediates (I) in Figures 1(a) and 1(b) are cutpoints. In Figure 1(c), I_1 and I_2 are cutpoints, but I_3 is not one. There are no cutpoints in Figure 1(d). Furthermore, $p_{pd}(S, T) = 1$ for Figures 1(a), 1(b), and 1(c), but $p_{pd}(S, T) = 2$ for Figure 1(d), which is consequently the best plan of these four.

The lines or the points of the synthesis digraph can be assigned *weights* according to, *e.g.*, the estimated yields of reactions based on literature precedents, the calculated strain energies of molecules, or topological indices.^{1,2,5}

A situation that occurs frequently in the synthesis of complex molecules is the production of a mixture of two (or more) products, only one of which can be converted into the desired product. This situation can also be represented by Figure 1(c), where a reaction performed on I_1 yields two products, I_2 and I_3 , and only I_2 can be converted directly into T. The overall yield can be improved if the wrong product can be converted into the correct one, either directly ($I_3 \rightarrow I_2$) or *via* conversion back to the starting material ($I_3 \rightarrow I_1$) followed by recycling. Both of these situations arose in Schreiber's recently completed synthesis of invictolide.⁶ In Figure 1(d) if I_1 yields a mixture of I_2 and I_3 , the possibility exists that each of them can be converted into T. If both conversions can be carried out simultaneously (in the same flask), the plan becomes reflexive.⁷

It may be conjectured that as the complexity of synthetic targets increases, corrective steps will become the rule, rather than the exception, since selectivity cannot be expected to increase proportionately with complexity. In fact, selectivity tends to decrease with increased complexity owing to the increased probability of side-reactions caused by interfering functionality.^{2,5} Developments in synthetic methodology will undoubtedly ameliorate this problem, but they cannot be counted upon to eliminate it. For example, a number of unexpected side reactions in complex intermediates delayed the synthesis of dodecahedrane for years until alternative routes were developed.⁸

While it is possible to inspect simple synthesis plans visually, and while some of the concepts introduced here are quite intuitive, computerized algorithms^{4,9} will be required in order to analyse and to evaluate complex synthesis plans, and these algorithms will have to be based upon the mathematically rigorous definition of concepts such as those introduced here.

Received, 23rd June 1986; Com. 856

References

- 1 E. J. Corey and W. T. Wipke, *Science*, 1969, **166**, 178.
- 2 J. B. Hendrickson, *J. Am. Chem. Soc.*, 1977, **99**, 5439.
- 3 (a) F. Harary, 'Graph Theory,' Addison-Wesley, Reading, Ma., 1969. This is a good introduction to graph theory. For a deeper discussion of digraphs see (b) F. Harary, R. Z. Norman, and D. Cartwright, 'Structural Models: An Introduction to the Theory of Directed Graphs,' Wiley, New York, 1965.
- 4 S. Even, 'Graph Algorithms,' Computer Science Press, Potomac, Md., 1979.
- 5 S. H. Bertz, *J. Am. Chem. Soc.*, 1982, **104**, 5801; S. H. Bertz, in 'Chemical Applications of Topology and Graph Theory,' Studies in Physical and Theoretical Chemistry, vol. 28, ed. R. B. King, Elsevier, Amsterdam, 1983, p. 206.
- 6 S. L. Schreiber and Z. Wang, *J. Am. Chem. Soc.*, 1985, **107**, 5303.
- 7 S. H. Bertz, *J. Chem. Soc., Chem. Commun.*, 1984, 218.
- 8 R. J. Ternansky, D. W. Balogh, and L. A. Paquette, *J. Am. Chem. Soc.*, 1982, **104**, 4503; D. W. Balogh, L. A. Paquette, P. Engel, and J. F. Blount, *ibid.*, 1981, **103**, 226; D. W. Balogh and L. A. Paquette, *J. Org. Chem.*, 1980, **45**, 3038; L. A. Paquette, *Pure and Appl. Chem.*, 1978, **50**, 1291.
- 9 M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Computers & Chemistry*, 1979, **3**, 5; 1980, **4**, 27.