THE GROWTH OF A WORD NETWORK

A. ROSS ECKLER
Morristown, New Jersey
LEONARD J. GORDON
Chico, California

The May 1973 issue of Word Ways introduced the concept of a word network, a connected diagram of words of a single length, in which any two words differing in only one letter in the same position (as BRA and ERA, or TON and TEN) are joined by a line. Networks with more than 50 to 100 words are generally too complicated to diagram on a single sheet of paper, and it is difficult to assess their properties.

In particular, it is hard to assess the degree of connectivity of a network—is any pair of words connected by many different routes, or are there isthmuses (such as Panama between North and South America) through which most routes must go? If most of the words are contained in a single main network, with only a few words or small groups of words left out, it is conjectured that no major isthmuses exist. However, it is possible that such isthmuses temporarily form as networks are built up a word at a time, joining formerly isolated groups for the first time.

To evaluate this possibility, this article shows how one can chart the evolution of a word network from a set of unconnected words. The diagram on the next two pages shows how it can be done, using three-letter words as an example. Kucera and Francis's Computational Analysis of Present-Day American English (Brown University Press, 1967) tabulates the observed frequencies of words in a million-word sample of materials printed in the United States in 1962. One can build up a network of words one at a time in decreasing order of frequency, starting with THE, AND, WAS, FOR, and HIS. At first, no words are connected. The initial link occurs between HIS (ranked 5th in the list) and HIM (ranked 15th). When HAS (6) appears, it joins the HIS-HIM fragment with two isolated words, HAD (6) and WAS (3). All this is diagrammed at the top of the tree-like structure on the next page, in which each word is followed by its Kucera-Francis ranking. The only word that might be fairly identified as an isthmus is NOR (69), which unites the 18-word network on the left with the 14-word network on the right by joining FOR in the former with NOT (or NOW) in the latter. (The parenthesized number preceding NOR indicates that the main network now has 33 members, or 48 per cent of all the words thus far sampled: 33/69 = 0.48.) Note that the isthmus begins to disappear with the very next word added, LOW, which joins LAW with NOW (or HOW).

Scanning the diagram, one sees that the main network soon becomes so large that it attracts other word groups to it before they
concept of a single length, in the same way by a line. It is difficult to connectivity any different en North and most of the only a few postulated that such isthmus a at a time, shen can chart on the right hand unites the law with. (The main network of words thus dissolves LAW with)

work soon begins before they
become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor

become ver-
dimensional
ably before
to be in
joining wor-

By the
abbreviati-
a number
ALL-ILL, is
reached,

Each pa-
a series of
never, then
all word-pa-
of these
network. In

can-m-
the span is

It is in
evolves. Or
imum value
out of isla-
merely cre-
work grows
if an island
nula is for-

In the
a 14-word
island with
with a span
the-she-s

There are
Dictionary
span of the

ivy-icy-ic

There are
Dictionary
and OXY—a
duced to

ivy-icy-ic
ivy-icy-ic

Some paths
are 33 wor
become very large. There are no hard-to-reach places in three-di­

men sional word space where a group of words can build up considerably before joining the main network. The greatest potential seems to be in words beginning with a vowel, which have difficulty in joining words beginning with a consonant.

By the time one has sampled 169 Kucera-Franics words (omitting abbreviations such as REV(erend), SEN(ator), and AUG(ust)), the main network has incorporated 85 per cent of the sample, leaving a number of isolated words plus the groups WHO-WHY, HER-PER, ALL-ILL, and TWO-TOO-TOP-TOM. When the next word, TIM 170, is reached, this four-word group is brought into the main network.

Each pair of non-adjacent words in a network can be joined by a series of intermediate words, often in more than one way; however, there is always some path of minimum length. Looking at all word-pairs in the network, the number of steps in the longest of these minimum-length paths is defined to be the span of the network. In the simple network

\[
\begin{align*}
\text{had} & \\
\text{can-man-may-way-was-has-his-him} & \\
\text{men} &
\end{align*}
\]
the span is 7, achieved by CAN (or MEN) to HIM.

It is instructive to see how the span changes as the network evolves. One expects it to increase with network size to some maximum value, probably achieved when the main network first forms out of islands, and then slowly decrease when there are no large islands left to annex and most of the words added to the network merely create alternate paths and short-cuts. However, as the network grows, the span may temporarily increase. This will occur if an island being annexed is moderately large, and the point of annexation is near the edge of the main network, so that a peninsula is formed.

In the evolutionary network diagrammed previously, NOR joins a 14-word island with a span of 9 (THE to FEW) to an 18-word island with a span of 7 (CAN to HIM), to form a 33-word network with a span of 14:

\[
\begin{align*}
\text{the-she-see-set-get-got-not-nor-for-far-war-was-has-his-him} & \\
\text{ivy-icy-ice-ire-ere-err-ear-bar-bay-say-sky-ski} &
\end{align*}
\]

There are 539 three-letter words in the Merriam-Webster Pocket Dictionary (excluding abbreviations such as DDT and TNT); the span of the main network has been reduced to 11:

\[
\begin{align*}
\text{ivy-icy-ice-ace-aye-tye-the-thy-try-fry-fro} & \\
\text{ivy-icy-ice-aye-tye-the-thy-try-pry-pro} &
\end{align*}
\]

Some paths of the network are extremely dense; for example, there are 33 words one step away from PAT, and 244 more two steps away.
Probably one-third of the network lies within two steps of the cycle PAT-PAY-SAT-SAY.

How do corresponding networks of four- and five-letter words evolve? Their behavior is very similar to the three-letter one: at first a large number of islets which coalesce into a few large islands, then a grand coalition of the major islands into a main network containing some 40 to 50 per cent of the words in the sample. However, these newly-formed main networks are much larger in size. The four-letter main network is formed when HOLE unites a 71-word island with a 56-word one, followed immediately by LOSE which unites the resultant 128-word island with a 27-word one. The newly-formed main network of 156 words uses 47 per cent of the 331 words sampled to that point.

As the four-letter main network is five times as large as the three-letter one, it is not practicable to show the detailed diagram of its growth here. However, one can capture the flavor of the network growth by noting those words which unite the network with islands. For example, in the three-letter network, WAY joins an island of size 5 containing WAS, to an island of size 3 containing MAY, to form a new island of size 9. By the time this has grown to size 13 by four single-word accretions (not specified below), the word FAR links WAR with the single-word island FOR to form a new island of size 15. The growth of the largest island is thus charted until it becomes so overwhelmingly larger than other islands that it can be fairly termed the main network (at NOR).

5 has (his 2, was 1, had 1)
9 way (was 5, may 3)
15 far (war 13, for 1)
33 NOR (for 18, not 14)
42 ran (man 40, run 1)
44 led (let 42, red 1)
59 bit (hit 52, big 1, but 5)

76 die (did 74, due 1)
80 buy (but 77, boy 2)
83 dry (day 81, try 1)
103 odd (add 101, old 1)
109 Bob (boy 106, job 2)
119 aim (aid 111, am 7)
143 fog (fig 141, dog 1)
148 Tim (aim 143, Tom 4)

The corresponding list of annexations for the four-letter evolutionary network is given below. By the time the main network attains a size of 344 with FRED, it contains 67 per cent of the 513 sampled words.

4 gave (give 2, have 1)
11 live (five 5, love 1, like 4)
13 move (love 11, more 1)
17 fine (five 13, find 3)
27 give (gave 17, same 9)
31 firm (fire 28, form 2)
37 Mike (like 33, make 3)
44 nine (line 39, none 4)
52 wine (line 48, wide 3)

201 meat (meet 101, eat 1)
204 sale (same 104, eat 4)
222 seed (need 222, feed 1)
257 fail (fall 128, fail 2)
268 fool (foot 268, fool 1)
273 food (food 273, foot 2)
277 cash (case 233, cash 1)

The network
3 held (head 131, hold 6)
7 hold (held 64, hold 1)
10 read (head 103, read 1)
13 hear (head 97, hear 1)
31 hell (held 8, hell 2)
38 fell (hell 40, fell 1)
51 beat (heat 27, beat 5)
56 text (text 56, text 3)

There are many later.

what-that

which joins is sampled,

view-Viet

which joins

At the time

and lose-lost,

wait-want

However, the time when

293:

very-var

Curiously, the peninsula was

fall-fail

by the time

Curiously, the penin

The five one, again

main network

a 233-word
If the cycle letter words letter one: at large into a main in the sam-

HOLE unites 7-word one.

per cent of

and is thus

other IS-

Y joins an

FOR to form

which joins to SHiP in the main network when CHiP, the 636th word, is sampled, and

view-Viet-diet/died-tied
dies=ties=lies=lips
goes=does=toes
dogs

which joins to LOTS through DOTS, the 784th word in the sample.

At the time the main network is first formed with the aid of HOLE and LOSE, the span reaches a value of 23:

wait-want-went-west-best-beat-head-held-hole-role-rose-

lose-lost-last-fast-fact-pact-part-park-mark-Mary-

many

However, the largest known span of 25 is achieved for the first time when VERY is added to the network and it reaches a size of 293:

very-very-Mary-mark-park-part-pact-pace-pale-male-mile-file-fill-

fall-fail-fair-pair-paid-laid-land-band-bank-back-lack-luck-Lucy

Curiously, the 16-word island annexed by FLEW does not form a peninsula which extends span beyond 25. The span decreases to 14 by the time 3670 words in the OSPD have been sampled, as reported in the February 1989 issue of Word Ways.

The five-letter evolutionary network resembles the four-letter one, again scaled up by a factor of five or so. The five-letter main network is formed when BEATS unites a 383-word island with a 233-word one, BLINK unites this 617-word island with a 69-word
one, and BLOTS unites this 687-word island with a 27-word one, to form a main network with 715 words, 43 per cent of the 1664 words sampled to that point. The full list of annexations is:

5 lives (gives 2, lines 1, lived 1) 284 casts (cases 262, costs 21)
8 loved (lived 6, moved 1) 305 baths (Bates 303, baths 1)
18 mines (minds 9, moves 4, miles 4) 357 roofs (roots 355, roofs 1)
21 wines (wines 19, waves 1) 359 rains (gains 349, reins 3)
28 wives (lives 23, waves 4) 387 beaks (backs 375, beaks 1)
33 lover (loved 29, love 1) 389 Kenny (Kenny 388, Kenny 1)
41 males (miles 34, sales 6) 401 hands (hands 399, hands 2)
44 talks (tales 42, tasks 1) 404 banks (banks 402, banks 1)
47 tanks (talks 44, banks 2) 407 series (series 406, series 1)
54 liver (lived 56, river 1) 412 cities (cities 410, cities 1)
60 rider (river 58, wider 1) 414 hopes (hopes 412, hopes 1)
70 lever (liver 64, level 1, never 4) 416 issues (issues 414, issues 1)
87 files (miles 72, filed 14) 419 four (four 418, four 1)
89 Sally (Wally 87, silly 1) 422 only (only 420, only 1)
96 maker (makes 94, Baker 1) 425 longer (longer 423, longer 1)
98 sadly (sally 96, badly 1) 426 nuts (nuts 424, nuts 1)
100 tires (fries 98, times 1) 430 cans (cans 428, cans 1)
123 bands (banks 100, bonds 5, hands 17) 432 old (old 430, old 1)
126 wired (tired 124, wiped 1) 436 tops (tires 434, tops 1)
147 mates (males 127, rates 19) 440 lives (lives 438, lives 1)
155 rises (rider 148, rises 1, sides 5) 442 lures (lures 441, lures 1)
160 sings (wings 158, songs 1) 444 times (times 442, times 1)
172 caries (cared 162, cards 2, cases 7) 446 sides (sides 445, sides 1)
175 fully (Billy 172, fully 2) 448 ants (ants 447, ants 1)
180 pants (parts 178, wants 1) 450 books (books 448, books 1)
187 timed (times 182, aimed 4) 452 shoes (shoes 450, shoes 1)
191 belts (bells 189, belts 1) 454 pies (pies 452, pies 1)
210 doses (doses 193, horse 16) 456 cases (cases 454, cases 1)
216 merry (merry 213, mercy 1, Jerry 3) 458 likes (likes 456, likes 1)
221 modes (moves 219, codes 1) 460 books (books 458, books 1)
224 model (modes 221, motel 2) 462 beds (beds 460, beds 1)
233 forks (forms 231, folks 1) 464 cakes (cakes 462, cakes 1)
239 mails (mails 237, mails 1) 466 teeth (teeth 464, teeth 1)
246 molds (holds 239, moos 6) 468 days (days 466, days 1)
249 paces (faces 247, paced 1) 470 days (days 468, days 1)
252 tunes ( tones 249, dunes 1, tubes 1) 472 days (days 470, days 1)

The largest island to join the main network later contains 27 words:

fetch=letch
 latch=match=catch=watch=patch=batch=atch
 march=itch=itch=bitch=itch
 marsh=itch=Dutch
 harsh=punch=unch=unch
 bench=Reich
 beach=each=reach=react

This island joins to PEACE in the main network via PEACH.
By the time BEAME (the surname of a former New York City mayor) is reached, the main network still contains less than half of the words sampled (46 per cent).

At the time the main network is formed with the aid of BEATS, BLINK, and BLOTS, the span reaches a probable maximum value of 52:


The span is 48 at the time BEAME is added to the network. As reported in the February 1989 *Word Ways*, the span decreases to 29 for the 8200+ five-letter words in the OSPD.

Here is a summary of the span behavior in the evolutionary network as a function of word length:

<table>
<thead>
<tr>
<th>Word Length</th>
<th>Span at Time of Coalition</th>
<th>Largest-Known Value of Span</th>
<th>Ratio of Largest to OSPD Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>14</td>
<td>17</td>
<td>1.70</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>25</td>
<td>1.79</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>29</td>
<td>1.79</td>
</tr>
</tbody>
</table>

As the word length increases, the span at the time of coalition seems to approach the maximum span in value.

Finally, one should note that the Kucera-Francis ranking of words in decreasing order of frequency is to a considerable extent dependent upon the sample of words drawn -- a different million-word sample would change the relative ranking of all but the few dozen commonest words. There is no particular significance in the fact that NOR, HOLE, LOSE, BEATS, BLINK, and BLOTS were the essential links that first formed the main networks; in another sample, other words would assume this role. Nevertheless, it is likely that the statistical features of this study -- the size of the maximum span, the percentage of the sampled words included in the main network when it first forms, etc. -- would be approximately the same if a new sample were taken. In fact, one might even be able to simulate the behavior of evolutionary word networks by taking random samples of various sizes from the OSPD word lists. This is, in fact, the only way available to study evolutionary networks with words of six letters or more, for the Kucera-Francis sample is not large enough even to show the coalition of islands into a main network.