

지적 구조의 시각화를 위한 네트워크 형성 방식에 관한 연구

A Study on the Network Generation Methods for Visualizing
Knowledge Domain's Intellectual Structure

이재윤, 경기대학교 {memexlee@kgu.ac.kr}

Lee, Jae-Yun, Kyonggi University

지적 구조 분석을 위해서 계량서지적 자료를 시각적으로 표현하는 다양한 네트워크 형성 방식에 대해서 사례와 함께 각각의 특성을 살펴보았다. 동시인용을 비롯한 계량서지적 자료의 시각적인 표현은 지적구조의 분석에 있어서 매우 효과적인 방법으로 인식되어왔다. 시각화를 위해서는 전통적으로 다차원척도법이나 군집분석, 대응일치분석 등의 다변량통계 분석 기법이 사용되어왔다. 최근 들어 패스파인더 네트워크를 비롯한 새로운 네트워크 형성 기법이 전통적인 기법의 대체 도구로 제시되고 있으나, 사실상 네트워크의 형태로 지적 구조를 분석하는 것은 인용분석 연구의 초기부터라고 할 수 있다. 다양한 네트워크 형성 방식의 특성에 대해서 살펴봄으로써 계량서지적 분석을 활성화하는데 도움이 되리라고 기대한다.

1. 서론

이 연구의 목적은 학문 분야를 비롯한 특정 영역의 지적 구조를 시각적으로 표현하는 도구로서 네트워크 형성 방식에 대해 고찰하는 것이다.

계량서지학 분야에서 지적 구조의 시각화를 위해 사용되어온 전통적인 기법은 다차원척도법(MDS), 군집분석 등이다. 비록 분석 대상은 문헌 동시인용(Small 1973), 저자동시인용(White & Griffith 1981)을 비롯하여 저널동시인용, 동시출현 단어, 웹사이트 동시링크 등으로 변화해왔지만 시각화 방식은 크게 다르지 않았다.

네트워크 표현 방식은 동시인용분석을 제안한 Small(1973)도 사용했을 만큼 오래된 기법이다(<그림 2> 참조). 이 당시에는 개별 문헌네트워크와 문헌군집 네트워크를 별도로 그려서 지적 구조를 분석하였다(Garfield 1979). 그러나 저자동시인용분석을 제안한 White & Griffith(1981)가 사용한 다차원척도법과 군집분석은 개별 문헌·저자

와 군집을 동시에 표현하면서 전체적인 지적구조를 보여줄 수 있는 장점 때문에 이후 연구에서 널리 사용되었다. 상대적으로 이전의 단순한 네트워크 표현방식을 사용하는 연구는 줄어들었다.

최근에는 다차원척도법도 지적 구조의 시각화 도구로서 한계가 있음이 여러 연구에서 지적되고 있다. 관련 주장을 정리하면 다음과 같다.

우선 표현해야 할 개체가 너무 많으면 SPSS와 같은 통계패키지의 다차원척도분석 모듈이 수용할 수 없다(100개가 한계). 또한 수십 개 이상의 개체를 2차원 MDS 지도로 표현할 때 세부구조의 표현력이 매우 떨어진다(Schvaneveldt 1990; Noel et al. 2003). 표현된 차원을 직관적으로 해석하기가 어려울 뿐더러(Börner et al. 2003), MDS 지도만으로는 소주제 집단을 식별할 수 없다. 군집분석으로 보완하더라도 MDS 지도가 세부구조의 표현력이 떨어지므로 실제 가까운 노드와 지도상 가까운 노드가 일치하지 않아서 구불구불한 군집이 그려진다.

지적 구조의 시각화를 위한 도구로 네트워크 표현이 다시 각광을 받게 된 것은 인지심리학 분야에서 개발된 패스파인더 네트워크(Pathfinder Network; PFNET으로 약칭) 알고리즘(Schvaneveldt 1990)을 Chen(1999), White(2003) 등이 저자동시인 용분석에, Marion & McCain(2001)이 저널동시인 용에 도입하면서부터이다.

이 연구에서는 최근 다차원척도법의 한계가 지적되면서 그 대안으로 다시 부각되고 있는 네트워크 형성 방식에 대해서 사례와 함께 여러 방식의 특성을 살펴보기로 한다. 실제 네트워크 형성 사례로는 기존 연구의 네트워크 그림과 함께, 네트워크 표현을 사용하지 않은 연구의 자료를 네트워크로 표현하여 제시한다.

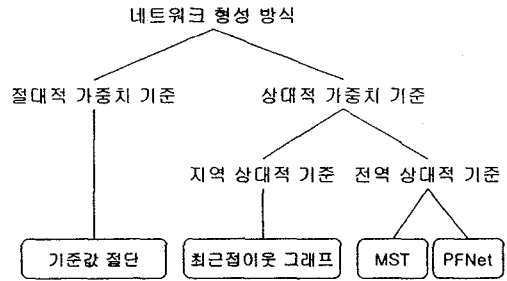
2. 네트워크 형성 방식

2.1 네트워크 형성 방식의 구분

주어진 연관성(또는 거리) 행렬로부터 네트워크를 형성하는 방식은 전체가 분리된 상태에서 가까운 노드 사이를 이어나가서 네트워크를 만드는 것이라고 할 수 있다. 이와 반대로 최대한의 링크가 연결되어 있는 상태에서 중요하지 않은 링크를 제거하는 것이라고 보는 입장에서는 링크 삭감 알고리즘(link reduction algorithm)이라고 부르기도 한다(Chen & Morris 2003).

어느 쪽으로 보든 간에 이 글에서 다루는 네트워크 형성 방식은 생성할(또는 제거할) 링크를 선택하는 기준에 따라 <그림 1>과 같이 구분할 수 있다.

주요한 링크를 선정할 때 각 링크의 가중치를 절대적 기준으로 사용하는 경우와 상대적 기준으로 사용하는 경우가 있을 수 있다. 기준값 절단 방식은 모든 링크의 가중치를 절대적인 기준으로 삼고 링크 생성 여부를 결정한다. 즉 생성된 링크의 가중치보다 가중치가 높으면서도 생략되는 링크는 있을 수 없다. 이와 달리 상대적 가중치 기준을 적용하는 방식에서는 가중치가 더 높으면서



<그림 1> 네트워크 형성 방식의 구분

도 상대적인 중요도에 따라서 생략되는 링크가 있을 수 있다. 최근접이웃 그래프는 각 노드별로 링크 가중치의 상대적인 높고 낮음을 기준으로 링크 가중치를 고려한다. 최소신장트리(Minimum Spanning Tree; MST로 약칭)나 PFNET은 각 노드별 상황과 전체적인 구조를 함께 고려해서 링크의 중요성을 상대적으로 판단한다.

다음 절부터는 각 방식의 특징에 대해서 사례와 함께 살펴보기로 한다.

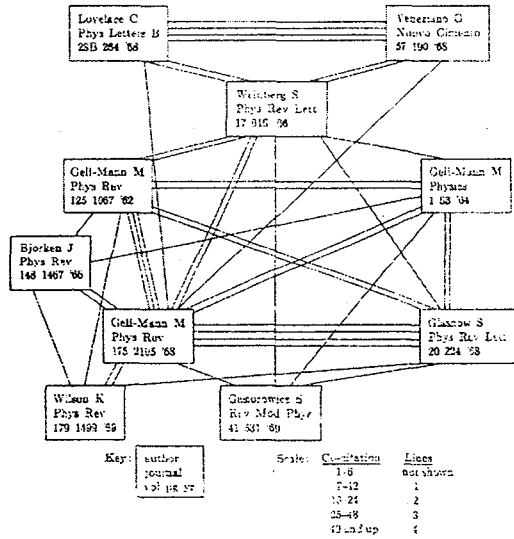
2.2 기준값 절단 방식

기준값 절단 방식은 가장 간단한 네트워크 생성 방식으로서 <그림 2>와 같이 동시인용네트워크 연구의 초기부터 적용된 방식이다.

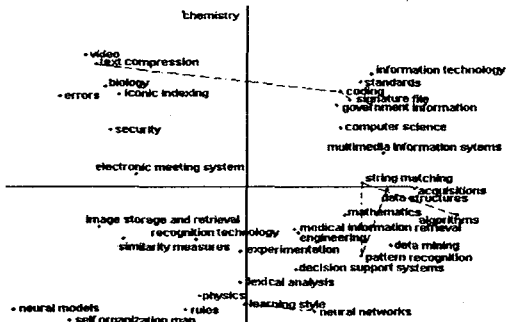
네트워크 전체적으로 가중치가 기준값 이상인 링크만 남기기 때문에 모든 링크에 대해서 절대적으로 동일한 기준을 적용하는 방식이다. <그림 2>의 경우에는 기준값이 7이며 이보다 값이 큰 경우에는 단계적으로 연결선의 수를 늘려서 강한 연결을 표현하였다.

이 방식으로는 기준값을 낮출수록 링크의 수가 많아져서 그래프가 복잡해진다. 따라서 핵심 저자나 주요 주제와 같은 정보가 부각되지 않는다. 반대로 기준값을 높이면 그래프가 분할되어 서브그래프나 고립 노드가 나타난다. 그 결과 전체적인 연결 흐름이 나타나지 않는 경우가 있다. 따라서 적당한 기준값을 찾는 것이 중요하다.

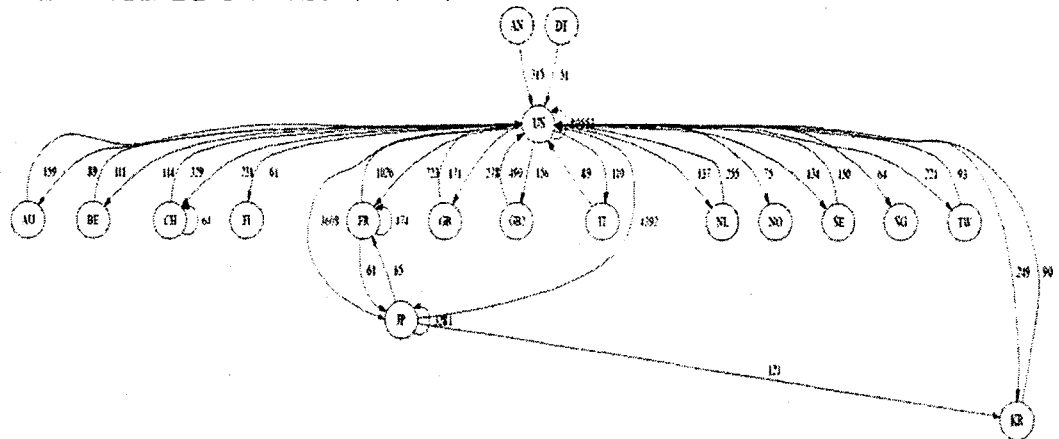
동시출현단어분석(co-word analysis)에서는 다차원척도법에 의한 MDS지도 상에서 보조 수단으로



<그림 2> 기준값 절단 방식 - 문헌동시인용 출처: Small(1973)



<그림 3> 기준값 절단 방식 - MDS 지도에 표시



<그림 4> 기준값 절단 방식 - 나노수준 과학기술분야 국가간 특허인용망 (빈도 50이상 연결 기준) 출처: Huang et al.(2003)

기준값 절단 방식을 사용하기도 한다(Ding et al. 2001). <그림 3>은 코사인유사도 0.2 이상인 키워드를 연결한 상태이다.

한편 공저자 네트워크 분석에는 무조건 한 편이라도 공저를 하면 링크를 생성하는 경우가 대부분이다(Newman 2001). 동시인용이나 단어동시출현에 비해서 공저라는 사건의 발생확률은 상당히 낮아서 1회라도 발생하는 것 자체가 학술 협력의 지표로서 의미가 있기 때문이다.

동시출현자료를 이용한 경우에는 대개는 고리(loop; 노드 자신에 대한 링크)가 없는 단순 그래프(simple graph)이지만 직접 인용망의 경우에는 드물게 <그림 4>처럼 고리를 표현한 경우도 있다.

2.2. 최근접이웃 그래프

최근접이웃 그래프(nearest neighbor graph; NNG로 약칭)는 각 노드마다 일정한 수(대개는 1 또는 작은 수 k)의 가까운 노드를 연결해서 생성된다(Eppstein et al. 1997). k를 붙여서 k-최근접이웃 그래프라고 하는 경우도 흔하다. 각 노드별로 상대적으로 가까운 노드를 연결하기 때문에 링크가 중치는 국지적인 관점에서 상대적인 기준으로 사용된다.

최근접이웃 그래프는 주로 사회과학 분야에서 설문이나 관찰을 통해서 인적 네트워크를 조사할

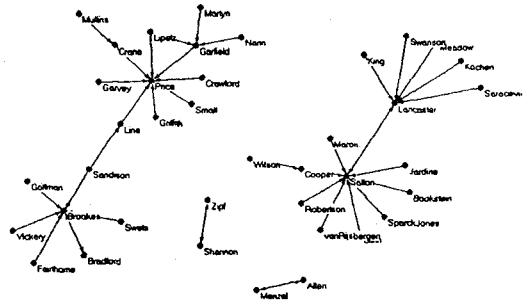
때 사용되어왔다(예: "17대 의원 네트워크 해부" - 조선일보 2004년 8월 24일자 A4-A6면). 그 이유는 모든 구성원 상호간의 관계를 설문이나 단기간의 관찰로 파악하는 것이 거의 불가능하기 때문이다.

하지만 최근접이웃 그래프는 몇 가지 장점을 가지고 있어서 동시인용 자료를 비롯한 계량서지 자료에 적용해볼 가치가 있다고 생각된다.

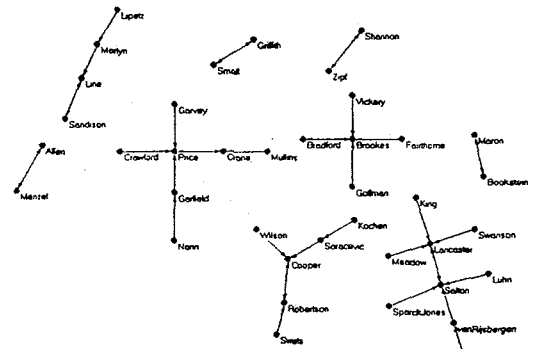
우선 최근접이웃 그래프에서는 기준값 절단 방식과 달리 다른 노드와 연결이 안되는 단독 고립 노드가 없다. 생성되는 가장 작은 단위는 노드 두 개짜리 서브그래프이다. k가 크면 전체가 하나의 그래프로 연결되지만 1이나 그에 가깝게 낮추면 여러 개의 서브 그래프로 분할되기 쉽다. 이 성질을 이용해서 군집을 생성할 수 있다. 그리고 각 노드마다 가까운 노드를 끊기 때문에 방향성 그래프를 생성할 수 있다. 따라서 지역 중심 노드의 파악이 손쉽다.

<그림 5>는 저자동시인용분석을 제안한 White & Griffith(1981)의 1970년대 정보학분야 저자동시인용자료를 입력해서 생성한 최근접이웃 그래프(k=1)이다. (a)는 동시인용빈도행렬을 그대로 입력한 경우이고, (b)는 코사인계수를 적용해서 동시인용빈도를 정규화한 행렬을 입력하여 처리한 결과이다. 빈도행렬을 입력자료로 하지 않고 코사인이나 자카드 계수로 정규화한 행렬을 대상으로 하는 경우에도 k값을 낮출 때와 마찬가지로 서브 그래프가 많아지는 경향이 있다. 이 경우에는 링크 가중치가 정규화되므로 빈도가 높은 특정 노드로의 집중경향이 약해지기 때문이다.

<그림 5>를 보면 빈도행렬을 이용한 (a)는 전체가 큰 두 개의 서브그래프(정보검색 및 정보시스템, 계량서지학 및 이론정보학)와 작은 서브 그래프 두 개로 구성된다. 반면에 코사인유사도행렬을 이용한 (b)는 9개의 중소규모 서브 그래프가 생성된다. 이를 White & Griffith(1981)가 실시한 요인분석 결과에 나타난 7개 요인과 비교해보면 거의 일치하였다. 따라서 최근접이웃 그래프는 소주제를 식별하는 유용한 도구가 될 가능성이 있다.



(a) 빈도행렬을 입력한 결과



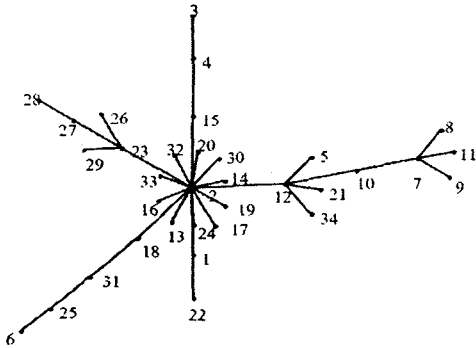
(b) 코사인유사도행렬을 입력한 결과

<그림 5> 최근접이웃 그래프(k=1) - 저자동시인용 저자동시인용자료 출처: White & Griffith(1981)

2.4 최소비용신장트리

어떠한 방식으로든 집합내의 모든 구성 노드를 일정한 과정을 거쳐서 차례대로 모두 연결시켜주는 방식을 신장트리라고 부른다. 흔히 그래프 탐색기법으로 불리는 깊이우선이나 너비우선 방식도 신장트리를 만드는 알고리즘이다. 신장트리 중에서도 모든 링크의 가중치 합이 가장 작은 트리를 최소(비용)신장트리(Minimum Spanning Tree: MST로 약칭)라고 부른다.

최소한의 비용으로 전체가 연결되도록 보장하기 위해서 링크 가중치가 아무리 높더라도 전체 구조상 중복된 링크이면 제거하는 상대적인 기준을 적용한다. 생성 결과는 노드 수가 n개인 경우 n-1개의 링크로 연결된 네트워크가 된다. 간단한 구현 알고리즘으로 Kruskal 알고리즘과 Prim 알고리즘이 널리 알려져 있다.

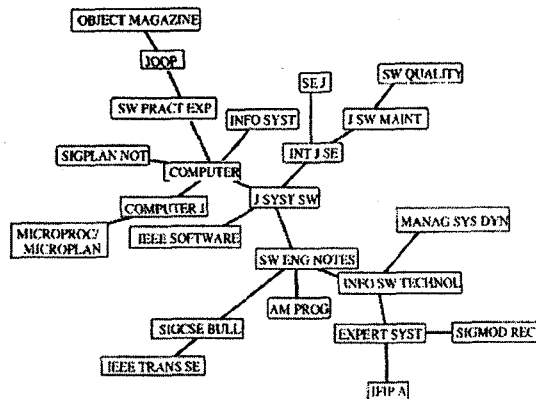


<그림 8> MST 사례 - 문헌동시인용
출처: Noel et al.(2002)

MST는 시각화 분야에서 흔히 사용되는 방식이지만 지적구조의 시각화를 위한 용도로 사용된 예는 Noel et al.(2002), Chen & Morris(2003) 정도에 불과하다. 세부구조의 표현능력 면에서 불리한 점이 있어서 다음 절에서 설명할 패스파인더 네트워크가 대신 사용되고 있다.

2.5 패스파인더 네트워크

패스파인더 네트워크(PFNET)는 원래 인지심리학 분야에서 개발하였으나 최근 계량서지적 네트워크 분석에서 널리 활용되고 있다. PFNET은 다차원척도법에 비해서 세부구조를 잘 표현할 뿐만 아니라 전체적인 구조도 뚜렷하게 제시해준다



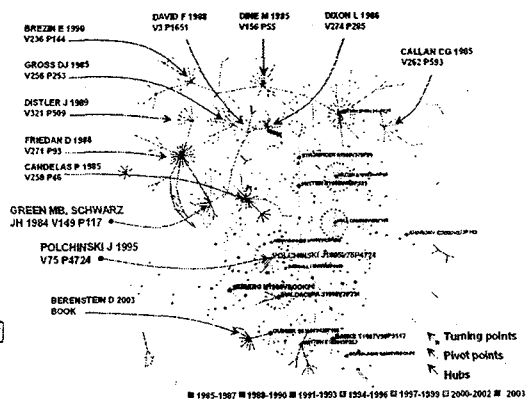
<그림 9> PFNET 사례 - 저널동시인용
출처: Marion & McCain(2001)

장점이 있다(Börner et al. 2003).

PFNET은 가중치가 있는 모든 링크가 생성된 상태(즉, 기준값 절단 방식에서 기준값을 0초과로 한 경우)에서 삼각부등식(triangle inequality)을 위반하는 경로를 제거하여 생성되는 네트워크이다 (Schvaneveldt 1990).

삼각부등식 위반 여부를 결정하기 위해서는 두 가지 파라미터 q 와 r 이 필요하다. 파라미터 q 는 노드 사이의 경로거리를 산출하는 데 고려하는 최대 링크의 수(거리산출범위)를 뜻한다. q 는 2에서 $n-1$ (n 은 노드의 총 수)까지 설정한다. q 가 커질수록 조사대상 범위가 넓어져서 엄격한 조건이 되므로 남는 링크의 수가 줄어든다. 파라미터 r 은 민코프스키 거리 공식의 제곱수로서 한 경로를 구성하는 여러 링크가 가지고 있는 가중치를 거리에 반영하는 방법을 뜻한다. r 이 1이면 각 링크 가중치의 합이 그대로 경로의 거리가 되고, r 이 무한대이면 경로를 구성하는 링크의 가중치 중 최대값이 경로의 거리가 된다. r 이 커질수록 경로의 길이가 짧아지므로 역시 엄격한 조건(위반되기 쉬운 조건)이 되어 남는 링크의 수가 줄어든다. 지적구조를 표현하기 위해서는 흔히 가장 엄격한 조건인 $r=\infty, q=n-1$ 로 설정(PFNET($r=\infty, q=n-1$)이라고 표기)하여 주요 흐름이 표현되도록 한다.

PFNET($r=\infty, q=n-1$)은 가능한 모든 MST를 합



<그림 10> PFNET 사례 - 문헌동시인용
출처: Chen(2004)

친 것과 같다(Schvaneveldt 1990). 즉, MST에서는 링크 가중치가 동일일 때 임의의 링크를 선택하는 대신 동일한 링크를 모두 선택하는 경우가 PFNET에 해당한다. 따라서 MST와 PFNET은 유사한 점이 많다.

특히 MST와 PFNET은 개별 노드의 전체적인 배치 구조면에서 상당히 비슷한 결과를 보여준다. 결과로 나타난 네트워크에 드러난 전체적인 흐름은 큰 차이가 없다. 그러나 세부 구조에 있어서는 PFNET이 MST보다 잘 표현해준다. MST가 동일한 링크 중 임의로 하나를 선택하기 때문에 큰 그림은 PFNET과 차이가 없어도 세부 연결정보에서 누락되는 부분이 발생한다.

반면에 PFNET이 MST에 비해서 알고리즘이 복잡하며 처리시간과 기억공간을 더 필요로 한다. 삼각부등식을 위반하는지 여부를 행렬 전체에 대해서 검사해야 하기 때문이다.

3. 행렬값 가공처리가 네트워크 구조에 미치는 영향

앞에서 살펴본 여러 네트워크 방식을 적용하기 위해서는 동시인용빈도와 같은 행렬 자료를 준비해서 입력해야 한다. 이때 행렬의 수치를 가공하는 방법에 따라서 생성되는 네트워크의 형태가 달라진다. 2장의 <그림 5>에 나타난 최근접이웃 그래프의 예를 보아도 빈도 행렬을 입력했을 경우와 코사인유사도로 정규화한 행렬을 입력한 경우에 다소 다른 구조가 만들어짐을 알 수 있다.

행렬값의 가공 방법에 따라서 지적 구조의 분석을 위한 입력 행렬은 1차 연관성 행렬과 2차 연관성 행렬로 나눌 수 있다.

1차 연관성 행렬은 빈도값을 그대로 이용하거나, 각자의 출현빈도(보정한 대각선값)를 감안하여 코사인계수 등으로 정규화한 값으로 구성된 행렬이다. 정규화 여부와 상관없이 1차 연관성 행렬은 양자간의 직접 동시출현 정도를 반영한 행렬이다.

2차 연관성 행렬은 White & Griffith(1981)가 제

안한 바와 같이 1차 연관성 행렬을 다시 가공하여 생성되는 연관성 행렬이다. 1차 연관성 행렬의 개별 프로파일 벡터간의 상관도를 피어슨 상관계수 등으로 한 차례 다시 계산하므로, 2차 연관성 행렬은 제삼자와의 동시출현 패턴의 유사함을 측정하는 행렬이다.

2차 연관성 행렬을 생성할 때 1차 연관성 행렬의 대각선값을 어떻게 보정하는가가 결과에 영향을 미친다. McCain(1990)과 같이 대각선값을 결측치로 처리하면 이에 상응하는 양자간의 동시출현 빈도가 연관성 산출에 전혀 반영되지 않기 때문에 순수한 2차 연관성을 측정하는 셈이다. 반면에 White & Griffith(1981)처럼 프로파일을 구성하는 값 중에서 가장 큰 값 셋을 더해서 2로 나눈 값으로 보정해서 사용하면 양자간의 직접 동시출현을 어느 정도 반영하는 2차 연관성을 산출하는 셈이다. 비유적으로 말하자면 순수한 1차 연관성과 순수한 2차 연관성의 중간이므로 1.5차 연관성이라고 할 수 있을 것이다.

연관성 산출 방식에 따른 행렬값의 차이를 확인하기 위해서 White & Griffith(1981)의 저자동시인용 자료로부터 여러 방법으로 가공한 수치로 구성된 행렬간의 상관도를 구해보면 <표 1>과 같다. 이 표를 보면 2차 연관성 중에서도 대각선을 보정한 경우가 결측으로 처리한 경우보다 원래의 빈도 행렬과 유사함을 알 수 있다.

Chen & Morris(2003)는 PFNET을 생성할 때에는 원 자료로 빈도행렬을 사용하는 것보다 자카드나 코사인과 같은 연관성척도로 정규화한 값을 사용해야 잠재적인 위험을 피할 수 있다고 지적

<표 1> 각 행렬간 상관도 (spearman 순위상관)

행렬을 구성하는 값의 유형	1차 연관성		2차 연관성 (상관계수)	
	빈도	코사인 유사도	대각선 보정	대각선 결측
1차 연관성 빈도		0.960	0.789	0.672
1차 연관성 코사인유사도	0.960		0.858	0.754
2차 연관성 대각선보정	0.789	0.858		0.977
2차 연관성 대각선결측	0.672	0.754	0.977	

하였다. 그 이유는 빈도행렬을 이용하면 동률값이 상대적으로 흔해서 링크가 더 많이 생성될 수 있으므로 최소한의 주요 링크만 남겨서 흐름을 볼 수 있다는 PFNET의 장점이 많이 희석되기 때문이다.

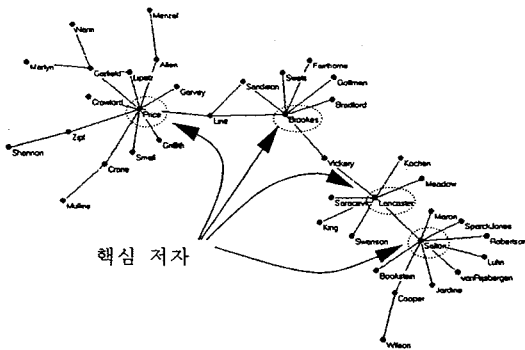
반면에 MST에 대해서 Noel et al.(2002)은 정규화한 값보다는 원 빈도 행렬을 입력해야 한다고 주장하였다. 원래의 값이 아닌 정규화한 값을 사용하면 많은 링크를 받는 핵심 노드가 부각되는 트리 구조가 생성되는 대신에, 노드가 줄지어 서는 사슬 효과(chain effect)가 나타나고, 그에 따라서 주요 노드와 관련 주제의 관찰이 어려워진다는 이유에서다.

사슬 효과의 문제는 PFNET에서도 지적된바 있다. White(2003)는 1차 연관성인 빈도행렬을 입력데이터로 사용하는 것이, 2차 연관성인 상관계

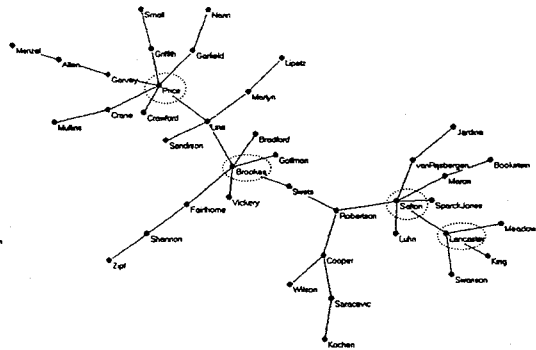
수행렬을 사용하는 것보다 바람직하다고 주장하였다. 그 이유는, 상관계수 행렬을 입력할 경우에 핵심저자를 부각시키는 스타노드(다수의 링크가 연결된 노드)가 사라지는 현상이 발생하기 때문이라고 하였다.

White(2003)의 주장에 대해서 유의할 부분은 상관계수 행렬을 산출하는 방법이다. 그가 2차 연관성을 산출할 때 사용한 방법은 대각선값 결측처리 방식이었는데 이것이 사슬 효과를 심화시킨 원인으로 여겨진다. 인용빈도가 높고 다수 저자와 동시인용되어 1차 연관성으로는 스타노드가 될 수 있는 저자가, 동시인용빈도의 높고 낮은 패턴까지 여러 저자들과 유사하지는 않기 때문에 순수한 2차 연관성 행렬에서는 스타노드가 되지 못하는 경우가 발생한다.

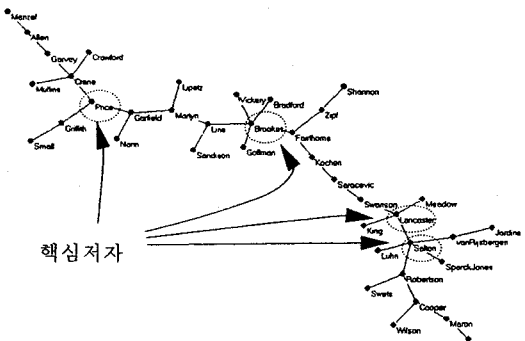
행렬값을 가공하는 것이 생성되는 네트워크에



(a) 빈도행렬로 작성한 PFNET



(b) 코사인유사도행렬로 작성한 PFNET



(c) 대각선값 보정후 상관행렬로 작성한 PFNET



(d) 대각선값 결측처리후 상관행렬로 작성한 PFNET

<그림 12> 입력행렬의 가공에 따른 PFNET의 사슬 효과 비교
저자동시인용자료 출처: White & Griffith(1981)

실제로 어떤 결과를 가져오는지 알아보기 위해서 <표 1>의 네 가지 행렬을 입력하여 생성된 네 가지 PFNET을 각각 <그림 11>에 제시하였다.

생성된 PFNET을 보면 빈도행렬을 입력한 경우인 (a)에서는 Price를 비롯한 네 명의 핵심저자가 많은 링크를 가지고 있어서 뚜렷하게 식별된다. 반면에 대각선값 결측처리후의 상관계수 행렬을 입력한 경우인 (d)는 사슬 효과가 심하게 나타나서 중심(스타)이라고 할 만한 노드를 찾기가 어렵다. 더군다나 핵심저자였던 Price 등은 간선이 아닌 가지의 끝에 위치하여서 별도로 부각되지 않는다. 같은 2차 연관성이라도 대각선값을 보정한 (c)에서는 사슬 효과는 다소 있지만 원래의 핵심저자가 간선 상에서 밀려나지는 않는다. 결국 대각선값을 결측처리하지 않는다면 White(2003)의 주장과 달리 2차 연관성 행렬로도 쓸모있는 PFNET 생성이 가능함을 알 수 있다.

결론적으로 빈도행렬을 입력하면 생성된 네트워크에서 소수의 저자만으로 큰 흐름을 볼 수 있다는 장점이 있다. 이와 달리 빈도값을 정규화한 행렬이나 대각선값 보정 후의 상관계수 행렬을 입력하면, 주제 면에서 중간적인 입장에 있는 비핵심저자들이 세부적으로 어떤 입지에 있는가를 살펴볼 수 있다.

6. 결론

패스파인더 네트워크를 비롯한 여러 네트워크 형성 기법의 특성에 대해서 살펴보았다. 여러 네트워크 형성 방식을 특성에 따라 적절히 사용할 수 있도록 함으로써 향후 계량서지적 분석을 활성화하는 데 도움이 되리라고 기대한다.

참고문헌

- Börner, K., C. Chen, and K. Boyack. 2003. "Visualizing knowledge domains." In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, 37, Medford, NJ: Information Today, Inc., chapter 5, pp. 179-255.
- Chen, C. 1999. "Visualising semantic spaces and author co-citation networks in digital libraries." *Information Processing & Management*, 35(3): 401-420.
- Chen, C. 2004. "Searching for intellectual turning points: Progressive knowledge domain visualization." *Proceedings of the National Academy of Sciences*, 2004, 101: 5303-5310.
- Chen, C., and S. Morris. 2003. "Visualizing evolving networks: Minimum spanning trees versus Pathfinder networks." *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'03)*. pp. 67-74.
- Ding, Y., G. G. Chowdhury, and Schubert Foo. 2001. "Bibliometric cartography of information retrieval research by using co-word analysis." *Information Processing & Management*, 37(6): 817-842.
- Eppstein, David, Michael S. Paterson, and Frances F. Yao. 1997. "On nearest neighbor graphs." *Discrete & Computational Geometry*, 17(3): 263-282.
- Garfield, E. 1979. *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley-Interscience.
- Huang, Z., H. Chen, A. Yip, G. Ng, F. Guo, Z.-K. Chen, and M. C. Roco. 2003. "Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field." *Journal of Nanoparticle Research*, 5(3/4): 333-363.
- Marion, L. S., and K. W. McCain. 2001. "Contrasting views of software engineering journals: Author cocitation choices and indexer vocabulary assignments." *Journal of the American Society for Information Science & Technology*, 52(4): 297-308.
- McCain, K. W. 1990. "Mapping authors in intellectual space: A technical overview." *Journal of the American Society for Information Science*, 41(6): 433-443.
- Newman, M. E. J. 2001. "Scientific collaboration networks. I: Network construction and fundamental results." *Physical Review E*, 64 p. 016131.
- Noel, S., C.-H. H. Chu, and V. Raghavan. 2003. "Co-citation count vs correlation for influence network visualization." *Information Visualization*, 2(3): 160-170.
- Schvaneveldt, R. W. (ed). 1990. *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.
- Small, H. 1973. "Co-citation in the scientific literature: A new measure of the relationship between publications." *Journal of the American Society for Information Science*, 24: 265-269.
- White, H. D. 2003. "Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists." *Journal of the American Society for Information Science & Technology*, 54(5): 423-434.
- White, H. D., and B. C. Griffith. 1981. "Author cocitation: A literature measure of intellectual structure." *Journal of the American Society for Information Science*, 32(3): 163-171.