

국내 학술지의 인용 네트워크 지수 측정*

Measuring the Prestige of Domestic Journals in Korean Journal Citation Network

이재윤, 경기대학교, memexlee@kgu.ac.kr

최선희, 한국과학기술정보연구원, sunny.choi@kisti.re.kr

Jae Yun Lee, Kyonggi University

Seon-Heui Choi, KISTI

최근 Web of Science에 도입된 Eigenfactor지수와 논문 영향력 지수(Article Influence Score), 그리고 Scopus에 도입된 SJR 지수는 구글의 PageRank 알고리즘과 같은 네트워크 분석 방식의 인용 지수이다. 국내 인용 색인 데이터베이스는 인용 링크가 외부로 향하는 비율과 자기 인용 비율이 높으므로 기존의 네트워크 인용 지수 산출 방식을 그대로 적용하기에는 어려움이 많다. 이 연구에서는 국내 인용색인DB에 대해서 대표적인 네트워크 인용 지수인 저널 페이지랭크를 시험적으로 측정해보고 국내 학술지의 상황을 고려한 개선방안을 모색하였다.

1. 서론

구글의 검색결과 순위 기법으로 사용되어 널리 알려진 페이지랭크(PageRank) 알고리즘은 인용하는 웹 사이트의 중요도를 반영하여 인용 받은 웹 사이트의 중요도를 산출하는 알고리즘이다.

원래의 페이지랭크 알고리즘은 한 사이트에서 다른 사이트로 향한 링크가 하나씩 있는 경우를 모델링한 것이었다. 그러나 Bollen 등(2006)은 링크가 여러 개인 경우, 즉 가중 네트워크를 대상으로 하는 지수로 개발된 가중 페이지랭크 알고리즘(weighted PageRank; PRw)을 저널 인용 네트워크에 적용하여 저널의 중요성을 측정하는 시도를 하였다.

그후 인용 데이터베이스의 양대 산맥인 WoS와 Scopus에 페이지랭크 방식의 인용 지수인 논문 영향력 지수(Article Influence Score;

AIS)와 SJR 지수(Scimago Journal Ranking Indicator)가 각각 새로 도입되었다.

원래 AIS 지수와 SJR 지수는 Thomson ISI나 Elsevier와는 별개의 연구 기관에서 개발된 지표이지만, 그 유용성이 인정되면서 해당 데이터베이스에 공식 지수로 도입되기에 이르렀다.

이들 페이지랭크 방식의 인용 지수에 대한 관심이 증가하면서 Franceschet(2010b)은 Eigenfactor 지수를 사용해야 하는 10가지 이유를 논문으로 제시하기까지 하였다. 10가지 이유 중에서 페이지랭크 방식의 지수에 공통적으로 적용되는 사항은 인용하는 저널의 중요도를 반영한다는 점, 인용하는 저널의 참고 문헌 수를 반영한다는 점, 그리고 인용 네트워크 전체를 활용한다는 점이다.

Bollen 등(2006)이 제안한 저널 페이지랭크 방식과 이후에 개발된 논문 영향력 지수, 그리고 SJR 지수의 특징을 간단히 비교하면 <표 1>과 같다.

한편 국내 인용 색인 데이터베이스는 인용 링크가 외부로 향하는 비율과 자기 인용 비율이 높으므로 페이지랭크와 같은 기존의 네트워크 인용 지수 산출 방식을 그대로 적용하기에는 어

* 본 연구는 행정안전부가 지원하는 한국정보화진흥원의 국가DB사업의 일환으로 한국과학기술정보연구원 연구원이 구축하여 제공하는 과학기술 참고문헌 인용 색인 DB정보를 이용하였음.

<표 1> 페이지랭크 방식의 저널 평가 지수 비교

지수 명칭	가중 페이지랭크(PRw)	SJR 지수	논문 영향력 지수(AIS)
관련 사이트	없음	scimagojr.com	eigenfactor.org
인용 DB	JCR	Scopus	JCR
저널 자기 인용	포함	총 인용의 1/3만 포함	배제
인용 범위	지난 2년	지난 3년	지난 5년
댐핑 팩터	? (0.85로 추정)	0.9	0.85
관련 지수	Y-Factor (PRw×IF)	PSJR (SJR×논문 수)	Eigenfactor (AIS×논문 수)

려움이 많다. 특히 <표 1>에서 보듯이 논문 영향력 지수와 SJR 지수에서는 자기 인용 빈도를 아예 배제하거나 전체 인용빈도의 1/3로 제한하고 있으나, 자기 인용 비율이 높은 국내 학술지에 이와 같은 방식을 그대로 적용하는 것이 타당한지는 아직 판단하기 어려운 상황이다. 따라서 이 연구에서는 국내 인용색인DB인 KSCI에 대해서 네트워크 인용 지수를 시험적으로 측정해보고 국내 학술지의 상황을 고려하여 자기 인용 비율을 조정하는 새로운 방식을 제안한다.

2. 국내 학술지에 대한 페이지랭크 방식 인용 지수 측정

2.1 KSCI에서의 저널 페이지랭크 측정 방법

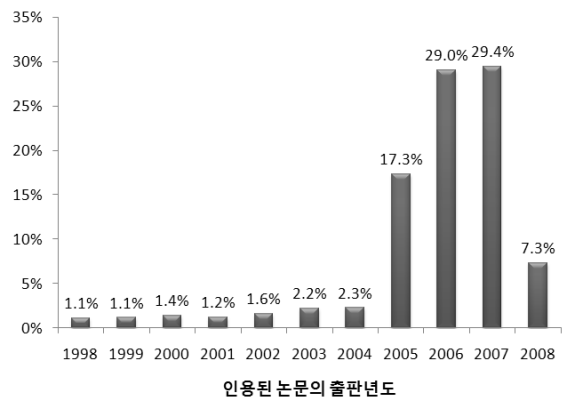
원래 KSCI 데이터베이스에는 452종의 저널이 공식적으로 등록되어 있으나, 데이터 분석 결과 2008년 인용 데이터와 2005년부터 2007년까지 3년간 발간된 논문 정보가 모두 포함된 저널은 449종으로 나타났다. 이 중에서 KSCI 데이터에서 발간 논문 수가 확인되지 않은 일부 저널은 해당 저널을 직접 확인하여 발간 논문수를 기록하였다.

인용 연도 범위를 판단하기 위해서 KSCI에 포함된 2008년 발간 학술지가 인용한 국내 저널 논문의 발간 연도를 살펴보면 <그림 1>과 같다. 이를 보면 2008년에는 이전 3년간인 2005년부터 2007년까지 발간된 논문이 주로 인용되고 있음을 알 수 있다. 이에 따라 인용 연도 범위는 SJR 지수와 마찬가지로 3년으로 설정하여 측정하였다.

2008년 KSCI 수록 논문에서 2005-2007년

KSCI 수록 저널의 논문 100,491건으로 향한 총 인용 빈도는 25,805건으로 파악되었다.

저널 페이지랭크 계산 공식은 SJR의 공식(González-Pereira et al. 2010)을 따랐으며, 각 저널의 중요도 변화가 수렴하는 기준은 Eigenfactor와 마찬가지로 0.00001 이하로 설정하였다. 참고 문헌이 파악되지 않은 저널의 중요도가 확산되는 비율인 댐핑 팩터(damping factor)는 SJR과 같은 0.9로 설정하였다. 각 저널의 자기 인용 비율 제한 방법은 Bollen 등(2006)과 같이 전혀 제한하지 않는 방법과, SJR에서처럼 총 인용의 1/3이 내로 제한하는 방법이외에 새로운 방안을 고안하였다. 자기 인용 비율 처리 방법에 대해서는 다음 절에서 상세히 다루었다.



<그림 1> 2008년 KSCI 저널의 참고문헌 발간 연도 비율

2.2 자기 인용 비율 제한 방법

Eigenfactor의 AIS 계산에서는 학술지의 자

기 인용을 0으로 처리하였다. 즉, 자기 인용을 완전히 배제한 것이다. 이와 달리 Scimago의 SJR 지수 계산에서는 학술지의 자기인용빈도를 총 참고문헌 수의 1/3로 제한하였다. SJR 방식에서는 타 저널 인용 비율이 2/3보다 크거나 같아야 하므로, 자기 인용 빈도는 다른 저널을 인용한 빈도의 1/2을 넘지 못하게 된다. 즉, 다른 저널을 인용한 빈도가 2일 때 자신을 인용한 빈도가 2일 경우에는 총 인용빈도가 4가 되어 자기 인용 비율이 2/4=1/2이 되어 1/3을 넘기게 되고, 자기 인용 빈도를 1로 제한하면 총 인용빈도가 3이 되어 자기 인용 비율이 1/3에 그치게 된다. SJR 방식의 자기 인용 비율 조정을 공식으로 표현하면 다음과 같다.

$$RSa_j = \min\left(RS_j, \frac{RO_j}{2}\right)$$

여기서 RSa_j 는 조정된 저널 j 의 자기인용 빈도이며, RS_j 는 저널 j 의 실제 자기인용 빈도, RO_j 는 저널 j 가 다른 저널을 인용한 빈도(총 참고문헌 수)이다.

그러나 자기 인용 비율이 대단히 높아서 1/3을 넘는 저널이 많은 국내의 인용 데이터베이스에서 SJR 방식을 쓸 경우에는 각 저널의 특성이 제대로 차별화되지 못한다는 문제점이 있다. 분석 대상인 KSCI 데이터에서 측정한 각 저널의 자기 인용 비율 평균은 51.4%에 달했다. 따라서 국내 저널에서 국내 저널은 자기 인용 비율이 1/3을 넘는 것은 예외적인 경우가 아닌 셈이다.

때로는 자기인용 빈도를 일률적으로 절반만 반영함으로써 해결하는 방식도 쓸 수 있으나, 자기인용 빈도가 높은 저널에 대해서는 그 절반이라도 자기인용 비율이 무척 높게 되므로 반복 계산으로 지수를 산출하는 PageRank 방식에서는 근본적인 해결책이 될 수 없다. 자기 인용을 절반으로 줄이는 경우는 공식으로 다음과 같이 표현할 수 있다.

$$RSa_j = \frac{RS_j}{2}$$

이런 문제점을 해결하기 위해서 이 연구에서는 자기 인용 빈도를 타 저널을 인용하는 비율에 비례하여 반영하는 새로운 방식인 CR1을 다음과 같이 제안한다.

$$RSa_j = RS_j \times \frac{RO_j}{REF_j}$$

이 방식을 사용하면 자기 인용 비율은 최대 50% 이내로 제한되며 실제 자기인용 비율의 정도에 따라서 조정된 자기인용 빈도가 선형적으로 달라지게 된다(<그림 2> 참조). 한편 자기 인용 비율을 CR1 방식으로 제한하면 원래 자기 인용 비율이 50% 이하로 낮은 경우에는 지나치게 자기 인용 빈도를 삭감함으로써 SJR보다 더 낮은 자기 인용 비율을 보이게 된다.

이런 단점을 없애기 위해 고안한 방식인 CR2는 다음과 같은 공식으로 조정된 자기 인용 빈도를 산출한다.

$$RSa_j = RS_j \times \left(1 - \left(\frac{RS_j}{REF_j}\right)^2\right)$$

CR2 방식으로는 자기 인용 비율은 총 참고문헌 수의 2/3, 즉 66.7%로 제한되며, 자기 인용 비율이 전체의 33.3%일때까지는 SJR 방식과 거의 유사한 비율로 조정된다(<그림 2> 참조).

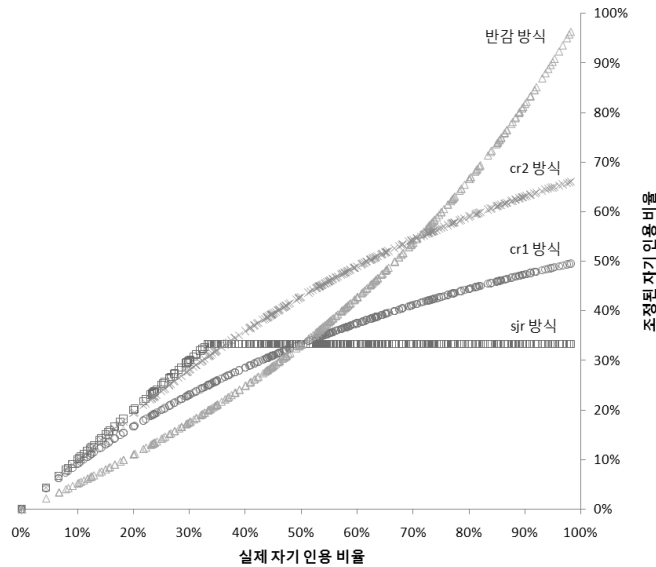
이상의 네 가지 자기 인용 비율 조정 방식을 KSCI에 등록된 저널에 적용하여 구한 자기 인용 비율의 분포는 <그림 2>와 같다. 이 중에서 SJR, CR1, CR2의 세 가지 방식에서는 원래 자기 인용 비율이 100%일때 자기 인용 빈도를 0으로 삭감하게 되며, 반감 방식에서는 원래 자기 인용 비율이 100%이면 조정된 자기 인용 비율도 100%가 된다.

3. 저널 페이지랭크 측정 결과 분석

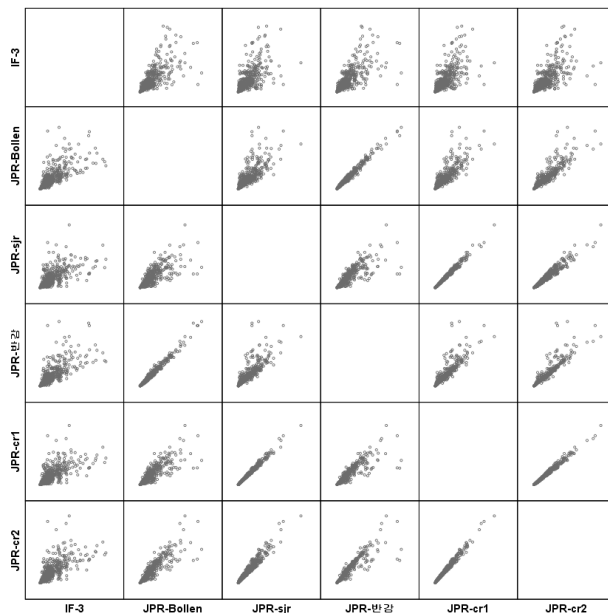
2005년부터 2007년까지 3년간 발간된 논문 정보와 2008년의 인용 정보가 확인된 저널 449종에 대해서 3년간 논문 수(ART), 참고문헌 수

(REF), 인용빈도(CIT), 자기 인용빈도(CIT-self), 자기 인용 비율(SR-rate), 타 저널에 대한 인용 빈도(REF-other) 등을 확인하였다. 저널 평가 지수로는 3년 단위 영향력 지수(IF-3)와 저널 페이지랭크를 측정하였으며, 저널 페이지랭크는 자기 인용 빈도를 Bollen 등(2006)과 같이 모두 포함한 경우(JPR-Bollen), SJR과 같이 총 인용의 1/3로 제한한 경우(JPR-sjr), 자기 인용의 절반

만 반영한 경우(JPR-반감), CR1과 CR2 공식에 따라 조정한 경우(JPR-cr1과 JPR-cr2)의 4가지 경우를 각각 측정하였다. 449종 저널에 대해서 측정된 5가지 인용 지수를 비교한 산포도는 <그림 3>과 같다. 또한 <표 2>는 F-3 지수를 기준으로 상위 10개 저널의 정보를 제시한 것이고, <표 3>은 JPR-cr2 지수를 기준으로 상위 10개 저널의 정보를 제시한 것이다.



<그림 2> 자기 인용 비율 조정 결과 비교



<그림 3> 449종 저널의 IF-3과 저널 페이지랭크 지수 비교 산포도

<표 2> IF-3 기준 상위 10개 저널의 각 지수별 순위와 통계량

저널명	지수별 순위						2008년 통계량 (논문 수는 '05-'07)					
	IF-3	JPR-bollen	JPR-sjr	JPR-반감	JPR-cr1	JPR-cr2	ART	REF	CIT	CIT-self	CO	SR-rate
한국식품조리과학회지	1	53	19	49	22	27	330	437	378	179	258	41.0%
Food Science and Biotechnology	2	37	56	45	41	36	544	614	612	458	156	74.6%
한국식품과학회지	3	29	4	18	5	6	430	191	470	64	127	33.5%
Biotechnology and Bioprocess Engineering:BBE	4	16	24	14	16	11	291	283	288	232	51	82.0%
대한지역사회영양학회지	5	46	27	42	23	23	210	181	204	90	91	49.7%
Journal of Microbiology and Biotechnology	6	9	45	10	31	22	812	701	781	627	74	89.4%
한국약용작물학회지	7	44	59	51	50	42	217	183	200	120	63	65.6%
Macromolecular Research	8	15	107	13	80	54	296	251	255	235	16	93.6%
설비공학논문집	9	3	122	4	83	53	388	303	333	293	10	96.7%
한국양봉학회지	10	4	121	2	86	63	71	54	60	53	1	98.1%

<표 3> JPR-cr2 기준 상위 10개 저널의 각 지수별 순위와 통계량

저널명	지수별 순위						2008년 통계량 (논문 수는 '05-'07)					
	IF-3	JPR-bollen	JPR-sjr	JPR-반감	JPR-cr1	JPR-cr2	ART	REF	CIT	CIT-self	CO	SR-rate
예방의학회지	34	6	1	5	1	1	197	55	100	27	28	49.1%
Korean Journal of Radiology	261	2	2	3	2	2	176	11	23	9	2	81.8%
Journal of Magnetics	46	12	5	7	3	3	115	41	52	37	4	90.2%
The Korean journal of hepatology	122	8	3	8	4	4	225	31	63	20	11	64.5%
국토계획	31	11	8	12	6	5	314	87	169	58	29	66.7%
한국식품과학회지	3	29	4	18	5	6	430	191	470	64	127	33.5%
International Journal of Automotive Technology	54	5	22	6	14	7	281	84	120	72	12	85.7%
분자세포학회지	63	10	18	9	10	8	337	101	135	86	15	85.1%
한국응용곤충학회지	24	17	11	17	9	9	159	95	97	62	33	65.3%
한국암석학회지	29	21	10	19	8	10	59	24	32	12	12	50.0%

<표 4> 학술지 통계량과 인용지수의 비교 - Spearman 순위상관

	ART	REF	CIT	CIT_self	SR_rate	REF_other	IF-3	JPR-Bollen	JPR-sjr	JPR-반감	JPR-cr1	JPR-cr2
ART	1.000	.677**	.734**	.674**	.309**	.512**	.211**	.340**	.251**	.312**	.270**	.297**
REF	.677**	1.000	.916**	.902**	.304**	.816**	.706**	.523**	.451**	.515**	.472**	.492**
CIT	.734**	.916**	1.000	.924**	.449**	.656**	.783**	.714**	.586**	.689**	.620**	.657**
CIT_self	.674**	.902**	.924**	1.000	.631**	.545**	.725**	.644**	.404**	.577**	.456**	.516**
SR_rate	.309**	.304**	.449**	.631**	1.000	-.199**	.410**	.571**	.101*	.432**	.191**	.297**
REF_other	.512**	.816**	.656**	.545**	-.199**	1.000	.493**	.220**	.440**	.281**	.406**	.368**
IF-3	.211**	.706**	.783**	.725**	.410**	.493**	1.000	.770**	.661**	.758**	.693**	.723**
JPR-Bollen	.340**	.523**	.714**	.644**	.571**	.220**	.770**	1.000	.806**	.981**	.861**	.911**
JPR-sjr	.251**	.451**	.586**	.404**	.101*	.440**	.661**	.806**	1.000	.888**	.993**	.973**
JPR-반감	.312**	.515**	.689**	.577**	.432**	.281**	.758**	.981**	.888**	1.000	.928**	.958**
JPR-cr1	.270**	.472**	.620**	.456**	.191**	.406**	.693**	.861**	.993**	.928**	1.000	.992**
JPR-cr2	.297**	.492**	.657**	.516**	.297**	.368**	.723**	.911**	.973**	.958**	.992**	1.000

** 99% 유의수준에서 유의함

* 95% 유의수준에서 유의함

<표 2>를 보면 IF-3 기준 상위 10개 저널 중에서 저널 페이지랭크 기준으로도 상위 10 이내에 포함되는 것은 지수별로 1~3종에 불과하였다. <표 3>에서 CR2 공식을 사용한 저널 페이지랭크 기준 상위 10개 저널 중에서 IF-3 기준으로도 10위 이내에 든 저널은 한국식품과학회지 1종에 불과했다.

<표 4>는 449종 학술지의 여러 통계량과 인용 지수 5가지 사이의 Spearman 순위 상관계수를 측정된 것이다. 이를 보면 IF-3와 페이지랭크 방식 지수의 순위 상관관계가 0.661에서 0.770 사이로 약 0.7 내외 정도로 나타났다.

Bollen 등(2006)의 연구에서는 IF와 자기인용을 그대로 두고 논문 수로 나누지 않은 저널 페이지랭크 값의 순위 상관관계수가 평균 0.61로 나타났으며, Franceschet(2010a)의 연구에서는 5년단위 IF와 자기인용을 배제한 Eigenfactor 지수와의 순위 상관관계수가 0.77, 논문 수로 나누기를 한 AIS와의 순위 상관관계수가 0.90로 나타났다. González-Pereira 등(2010)의 연구에서는 3년 단위 IF와 SJR 지수 사이의 순위 상관관계수가 27개 대분야별로 산출한 평균이 0.93으로 나타났다.

이를 보면 해외 데이터베이스에서는 논문 수로 나누지 않은 페이지랭크 방식 지수와 IF 사이의 순위 상관관계수가 0.61에서 0.77 사이로 나타나며, 논문 수로 나눈 경우에는 0.90 이상의 높은 순위 상관계수를 보였다. 반면에 KSCI 데이터베이스에서는 논문 수로 나눈 페이지랭크 방식 지수가 3년 단위 IF와 0.66에서 0.77 사이의 순위 상관계수를 보였으므로 IF와의 상관성이 상대적으로 낮았다. 이는 자기인용 비율이 상당히 높은 국내 학술지의 특성이 반영된 것으로 생각된다.

4. 결론

국내 인용색인 데이터베이스인 KSCI 수록 저널 중에서 449종을 대상으로 IF-3 지수와 다양한 저널 페이지랭크 지수를 측정된 결과,

순위에 상당한 차이가 있는 것으로 나타났다. 특히 해외 사례에 비해서 그 차이가 더 크게 나타난 것은 자기인용 비율이 높은 국내 저널의 상황 때문으로 짐작된다.

페이지랭크와 같은 네트워크 지수 방식의 인용 지수는 수록된 저널의 범위가 넓을수록 더 의미있는 지수 측정이 이루어진다. 네트워크 인용 지수에서는 직접 인용된 빈도가 바뀌지 않더라도, 한 저널의 중요도가 바뀌면 간접적으로 연결된 상당수의 저널의 중요도도 바뀌기 때문이다. 따라서 페이지랭크 방식의 인용 지수를 국내 저널에 적용한 결과가 더 의미를 가지기 위해서는 인용 데이터베이스의 범위가 더 포괄적이어야 하며, 논문 정보와 인용 정보의 파악도 더욱 철저하게 이루어져야 할 것이다.

참고문헌

- Bergstrom, C. T. 2007. "Eigenfactor: Measuring the value and prestige of scholarly journals." *C&RL News*, 68(5): 314-316.
- Bollen, Johan, Marko A. Rodriguez, and Herbert Van de Sompel. 2006. "Journal status." *Scientometrics*, 69(3): 669-687.
- Brin, S., L. Page, R. Motwani, and T. Winograd. 1999. The PageRank citation ranking: Bringing order to the Web. Technical Report 1999-66, Stanford InfoLab.
- Franceschet, M. 2010a. "Journal influence factors." *Journal of Informetrics*, 4(3): 239-248.
- Franceschet, M. 2010b. "Ten good reasons to use the Eigenfactor™ metrics." *Information Processing and Management*, 46(5): 555-558.
- González-Pereira, B., V. P. Guerrero-Bote, and F. Moya-Anegón. 2010. "A new approach to the metric of journals' scientific prestige: The SJR indicator." *Journal of Informetrics*, 4(3): 379-391.