

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФИЛОСОФСКИ ФАКУЛТЕТ
Книга Докторанти
Том 3, 2018

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF PHILOSOPHY
Postgraduate Students Book
Volume 3, 2018

НАУЧНИ ПУБЛИКАЦИИ СЪС ЗАБАВЕНО ПРИЗНАВАНЕ (SLEEPING BEAUTY): ЗНАЧЕНИЕ, МЕТОДИ ЗА ИНДЕНТИФИЦИРАНЕ И ОЦЕНЯВАНЕ

РОСЕН КРАСИМИРОВ СТЕФАНОВ

Анотация. Цитируемостта на дадена академична публикация често се приема като признак за нейното качество като научен труд. Прието е, че научни статии с голямо значение за научноизследователската общност привличат голямо внимание и цитати скоро след тяхното публикуване. Това твърдение често е валидно за статии, покриващи мащабни изследвания, които се публикуват в големи списания като Nature. В същото време съществуват статии, чиято научна стойност и значимост не бива подобаващо оценена при първоначалното си публикуване. Това са т.нар. „спящи красавици“ (sleeping beauties) на академичната среда. Този доклад представя същността на тези публикации, както и методите, които могат да се използват за оценяването им. Основният метод, разглеждан в този доклад, е В- (beauty) коефициента.

Разработката е резултат от изпълнение на проект „Проектиране и разработване на прототип на информационна система „Указател за цитируемост на публикации от български автори (обществени науки)“ (договор № ДН 15 /11.12.2017), финансиран от Фонд научни изследвания (МОН).

Ключови думи: спящи красавици, библиометрия, В-коефициент.

SCIENTIFIC PUBLICATIONS WITH SLEEPING BEAUTY: SIGNIFICANCE, METHODS OF IDENTIFICATION AND EVALUATION

Rosen Krassimirov Stefanov

Abstract. The quality of a given academic publication is often measured by how citable it is. A commonly accepted notion is that scientific articles of great scientific and research importance attract a large quantity of citations soon after they are published. This notion is often validated by articles that cover large research projects and which are published in large journals, such as Nature. Despite this there are articles whose academic importance is not properly recognized when they are initially published. These are the so called “sleeping beauties” of the academic en-

vironment. This report presents the nature of these publications, as well as the methods by which they can be evaluated. The main method researched in this report is the B- (beauty) coefficient.

The article is a result of the project “Design and Development of a Prototype of the Information System “Citation Index of Publications by Bulgarian Authors (Social Sciences)” (Contract № DN 15 / 11.12.2017), funded by the National Science Fund of Bulgaria (Ministry of Education and Science).

Key words: sleeping beauties, bibliometrics, B-coefficient.

* * *

Увод

Една от основните надежди на всеки научен изследовател е техният труд да бъде оценен подобаващо. Това означава признание на научната стойност на техните изследвания (и публикациите, в които те са вместени) и увеличаване на видимостта на тези изследвания в контекста на научноизследователската общност. Един от основните методи за измерване на това дали тези цели са били постигнати е това до каква степен научните публикации на изследователя се цитират. Това от своя страна допринася изследователите да искат да публикуват своите статии и други публикации в големи международни списания с висок импакт фактор. Желанието за признаване и повишена видимост на научните постижение не е основано само върху постигането на лично задоволство за добре извършено и полезно научно изследване.

В днешния академичен свят често се говори за предимствата на дигиталната революцията. За това как модерните информационни средства спомагат за по-ефикасното и бързо споделяне на информация, чрез каналите за електронно публикуване и модела open access. Ползата от тези средства е явна и целта на този доклад е не да оспори тяхната значимост. В същото време обаче, тази увеличена видимост за цялостната научноизследователската общност води до намаляване на шансовете на индивидуални изследователи или малки екипи да изпъкнат сред морето от научни публикации. В безбройните научни публикации индивидуалните трудове, независимо от тяхната научна стойност, често могат да останат незабелязани. Понякога тази липса на признание се дължи на чисто обективна причина: ниското качество на дадено научно изследване. Също така обаче съществуват и тези изследователи, чиито научни публикации представят съвсем реален принос за дадена научна среда, но не биват оценени подобаващо от научната общност при първоначалното им публикуване и съответно: не получават голям брой цитати. Това са научните публикации със забавено признаване, познати още като „спящи красавици“.

„Спящи красавици”: същност и примери

Преди да можем да опишем същността на „спящите красавици” на науката и да представим пример за такъв тип публикация, трябва да разгледаме как

протича нормалният „живот“ на една публикация от библиометрична гледна точка. При повечето научни публикации може да се наблюдава ясно изразен период, през който те получават голямо количество цитирания. След периода на висок интерес този брой цитирания постепенно намалява. Терминът описващ тази тенденция е *cited half-life* (от англ. ез.: период на полуразпад на получените цитирания). Предишни изследвания показват, че периодът на висока цитатна активност често се състои през първите 5 или по малко години, следващи първоначалното публикуване на дадена научна статия.¹ Естествено, тази тенденция е валидна само за „нормални“ научни публикации, които следват стандартния модел на цитатна активност. За публикации от типа „спяща красавица“, пикът на тяхната цитатна активност може да дойде десетилетия след тяхното първоначално публикуване. Именно това отклонение от нормата, въпреки безспорната им научна стойност, прави „спящите красавици“ изключително интересен обект за изследване.

Съществуването на публикации с късно разпозната научна стойност става обект на изследване през втората половина на 20. век. Първите изследвания върху този тип публикации са дело на Юджийн Гарфийлд и се фокусират върху възможността за използване на библиометрични системи за идентифицирането на такива изследвания.^{2,3}

За пръв път терминът „спящи красавици“ се използва за описването на недооценени научни публикации през 2004 г. в изследване на Антони ван Раан⁴. Подобно на едноименния разказ на братя Грим, „спящата красавица“ (научна публикация) потъва в дълбок сън (липса на цитатна активност), докато не бъде събудена от принц (публикацията или публикациите, които за пръв път започват да цитират неоцененото изследване). Но какви могат да бъдат причините дадена публикация да остане незабелязана с години, десетилетия или дори повече? Изследване на Реднер, фокусиращо се върху цитатни тенденции в полето на физиката, дава един възможен отговор на този въпрос. В изследването на Реднер могат да бъдат намерени публикации от едно общо научно поле, които получават значително внимание дълго след своето първоначално

¹ **Van Noorden**, Richard. Formula predicts research papers' future citations. *Nature*, 2013. <https://www.nature.com/news/formula-predicts-research-papers-future-citations-1.13881> (04.09.2018).

² **Garfield**, Eugene. Would Mendel's Work Have Been Ignored if Science Citation Index Was Available 100 Years Ago?. *Essays of an Information Scientist*, 1970, 1(2), pp. 69-70. <http://garfield.library.upenn.edu/essays/V1p069y1962-73.pdf> (04.09.2018).

³ **Garfield**, Eugene. Premature discovery or delayed recognition: why?. *Essays of an Information Scientist*, 1980, 4(21), pp. 5-10. <http://garfield.library.upenn.edu/essays/v4p488y1979-80.pdf> (04.09.2018).

⁴ **Van Raan**, Anthony. Sleeping beauties in science. *Scientometrics*, 2004, 4, pp. 461-466. <https://pdfs.semanticscholar.org/9acc/dbe65099e8be88ef5553e8833e98ba28250e.pdf> (04.09.2018).

публикуване.⁵ С други думи, един от факторите, които могат да имат влияние върху това дали дадена публикация е „спяща красавица“, е нейният тематичен обхват. Определени идеи и научни разкрития може да не бъдат приети от тогавашната научноизследователската общност поради дадена причина (прекалено специфичната природа на изследванията, например), но да намерят отзвук десетилетия по-късно, когато тези полета станат обект на завишен интерес.

При първоначалното си изследване на феномена „спяща красавица“, Ван Раан идентифицира три основни променливи, които определят същността на тези публикации⁶:

- Дълбочина на съня: определя се от броя на цитирания, които публикацията получава за всяка година на своя „сън“. За публикации с 0 или 1 цитирания се смята, че те са в „дълбок сън“. Публикации с 1 до 2 цитирания се намират в състояние на по „повърхностен сън“. Важно е да се отбележи, че при тези изчисления не се взимат предвид цитиранията, получени от други научни трудове на същия автор/колектив (самоцитирания).

- Продължителност на съня: колко години продължава горепредставеният процес.

- Интензивност на пробуждането: колко цитирания получава дадената публикация след своето пробуждане.

Един от най-известните примери за „спяща красавица“ в науката е публикацията „Може ли квантово-механичното описание на физическата реалност да се счита за завършено?“ с авторски колектив Алберт Айнщайн, Борис Подолски и Нейтън Росен.⁷ Основен научен продукт на изследването е Айнщайн-Подолски-Росен (АПР) парадокса, който се отнася до възможността за правенето на точни предсказания за една система от наблюденията върху изцяло друга и самостоятелна система. Спрямо данни от Google Scholar може да проследим цитатното развитие на тази публикация. За периода 1935–1965 (20 години след първоначалното публикуване) публикацията по квантова механика получава едва 77 цитирания.⁸ Десет години по-късно (1975) този брой вече е 201 цитирания, което се равнява на 161% ръст.⁹ През 2018 г. тази пуб-

⁵ **Redner**, Sidney. Citation Statistics from 110 years of Physics Review. *Physics Today*, 2005. <http://physics.bu.edu/~redner/pubs/pdf/PT.pdf> (04.09.2018)

⁶ **Van Raan**, Anthony. Sleeping beauties in science. *Scientometrics*, 2004, 4, pp. 461–466. <https://pdfs.semanticscholar.org/9acc/dbe65099e8be88ef5553e8833e98ba28250e.pdf> (04.09.2018).

⁷ **Einstein**, Albert et al. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?. *Physics Review*, 1935, 44 (777). <https://journals.aps.org/pr/abstract/10.1103/PhysRev.47.777> (04.09.2018).

⁸ “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935-1965. <https://scholar.google.bg/scholar> (04.09.2018).

⁹ “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935-1975. <https://scholar.google.bg/scholar> (04.09.2018).

ликация вече има над 15 000 цитирания.¹⁰ Интересно е да се отбележи, че през април 2018 г. АПР парадокса за пръв беше наблюдаван в комплексна система от частици (до този момент са провеждани експерименти между отделни атоми).¹¹ Това изследване показва как една идея, първоначално представена като „спяща красавица“ през първата половина на 20. Век, продължава да бива разглеждана и развивана 83 години по-късно.

Друг интересен пример за „спяща красавица“ е статията „Цитатни индекси в науката“ на Юджийн Гарфийлд, която може да се смята за основата за създаването на индекса за цитатен анализ „импакт фактор“ и системата Science Citation Index (днес част от Web Of Science)¹². Тази публикация не получава значително количество цитати (и съответно внимание) до края на 20. век. Точката на пробуждане за статията се смята през 2000 година, като нейният „принц“ е статия, оспорваща валидността на импактфактор-показателя като метод за оценяване на научната продукция.¹³ Интересното в този случай е, че публикацията на Гарфийлд се пробужда поради увеличаващата се към края на 20. век дискусията относно ролята на цитатните системи като метод за оценяване на качеството на световната научна продукция. По този начин, публикацията от средата на 20. век става обект на дискусията относно въпрос, който вълнува научната общност и до днес.

Методи за оценяване

През последното десетилетие са публикувани няколко изследвания, целящи да представят методи за идентифицирането на „спящи красавици“, както и за оценяването на „дълбочината“ на техния сън. Сред тях са изследванията на Ван Раан¹⁴ и Реднер¹⁵, но за целта на този доклад ще се фокусираме върху В- коефициента (beauty coefficient/V-coefficient) изготвен от Ке, Ферара,

¹⁰ “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935-2018. <https://scholar.google.bg/scholar> (04.09.2018).

¹¹ Einstein-Podolsky-Rosen paradox observed in many-particle system for the first time. University of Basel, 2018. <https://www.unibas.ch/en/News-Events/News/Uni-Research/Einstein-Podolsky-Rosen-paradox.html> (04.09.2018).

¹² **Garfield**, Eugene. Citation indexes for science: a new dimension in documentation. Science, 1955, 122, pp. 108–111. <http://garfield.library.upenn.edu/papers/science1955.pdf> (04.09.2018).

¹³ **Redner**, Sidney. Citation Statistics from 110 years of Physics Review. Physics Today, 2005. <http://physics.bu.edu/~redner/pubs/pdf/PT.pdf> (04.09.2018)

¹⁴ **Van Raan**, Anthony. Sleeping beauties in science. Scientometrics, 2004, 4, pp. 461-466. <https://pdfs.semanticscholar.org/9acc/dbe65099e8be88ef5553e8833e98ba28250e.pdf> (04.09.2018)

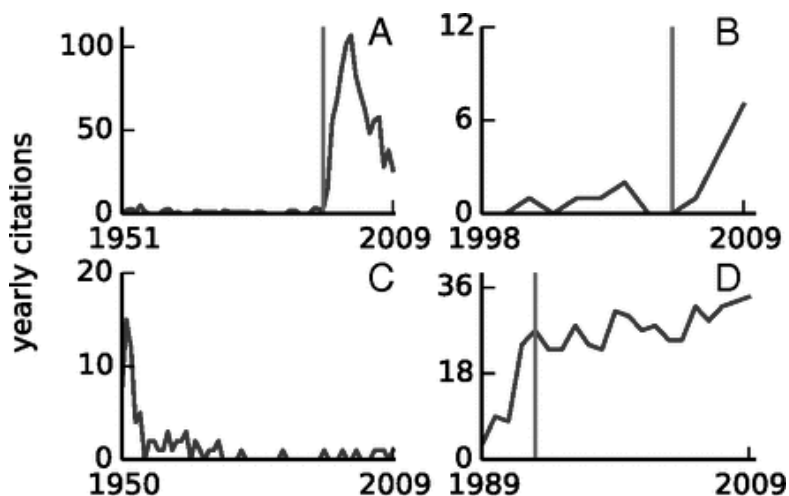
¹⁵ **Redner**, Sidney. Citation Statistics from 110 years of Physics Review. Physics Today, 2005. <http://physics.bu.edu/~redner/pubs/pdf/PT.pdf> (04.09.2018)

Радичи и Фламини.¹⁶ Изборът на това изследване като основен обект на този доклад се базира на два фактора. На първо място, изследването на Ке и неговите колеги покрива над 22 милиона публикации, вметени в базите данни на American Physical Society и Web of Science, покриващи хронологичен обхват от над 100 години (в сравнение, изследването на Ван Раан се ограничава до 20-годишен период). На второ място се отбелязва и полидисциплинарният обхват на изследването, благодарение на което може да се анализират „спящите красавици” от различни научни полета (за сравнение, изследванията на Ред-нер са фокусирани изцяло върху полето на физиката).

В-коэффициентът на Ке и неговите колеги е базиран на следните фактори:

- година на публикуване;
- продължителност на „съня”;
- максимален брой получени цитирания за всяка година;
- година на пробуждане и сила на пробуждането.

Последният фактор, момента на пробуждане, е от основно значение за определяне на финалния В-коэффициент, поставен на дадена статия. Той действа като разделяща линия между двете основни точки на анализ: (1) честотата на цитирания преди пробуждане и (2) честотата на цитирания след пробуждане. В разгледаното изследване е представена като червена линия (фиг. 1.)



Фиг. 1. Точка на пробуждане на публикации в изследването *Defining and Identifying Sleeping Beauties in Science*¹⁷

¹⁶ Ке, Qing et al. *Defining and identifying Sleeping Beauties in science*. PNAS, 2015, 112(24), pp. 7426–7431. <http://www.pnas.org/content/112/24/7426> (04.09.2018).

¹⁷ Пак там.

Публикация А има най-високият В-коэффициент (1722): наблюдава се дълъг период на „сън” и много интензивно пробуждане. Публикация В притежава по нисък В-коэффициент (22) поради по-голямото количество цитирания, получени преди пробуждането и по-ниската интензивност на това пробуждане. Публикация С има нулев В-коэффициент, защото следва стандартен модел на цитатна активност, където повечето цитирания към труда са получени в началото на неговия живот. На последно място, публикация D има постепенно повишаваща се тенденция относно цитируемостта, която спада в точката на пробуждане и съответно има негативен В-коэффициент (-5). Най-високият идентифициран В-коэффициент в изследването принадлежи на публикация в полето на физиката, публикувана през 1906 г. и пробудила се през 2002 (В-коэффициент=11 600). От основно значение за формирането на В-коэффициента не е толкова продължителността на „съня”, а интензивността на пробуждането. Друга изследвана публикация е издадена през 1901 г. и пробудена през 2002 г., но има по нисък В-коэффициент от този на гореспоменатата публикация (В-коэффициент=3978). Заключение, което може да се направи тук, е, че интензивността на пробуждането се приравнява на стойността на преоткритата научната стойност на публикацията.

Разпределението на спящи красавици по дисциплина е също разгледано. За целта Ке и колегите му анализират само публикации, вметени в базата данни на WoS. Това решение е напълно разбираемо, поради неравенството, което би се получило при включването на базата данни на APS (информационна система, фокусирана изцяло върху полето на физиката). Въпреки това, най-големият брой „спящи красавици” в WoS принадлежи към полето на физиката (7.6%), следвано от химията (7.5%) и математиката (4%).¹⁸

Полидисциплинарната същност на изследването може да се смята за основа за едно от най-важните му разкрития: „спящи красавици” от едно научно поле могат да бъдат пробудени от публикации, принадлежащи към изцяло друго поле. Пример за това е статия в полето на социалните науки на Уейн Закари, публикувана през 1977¹⁹. Статията се фокусира върху изследване на малка социална група (в този случай, Клуб по карате) и това как събития в рамките на тази група могат да доведат до драстични изменения в нейната структура. Тази публикация остава незабелязана до началото на 21. Век, докато Гирван и Нюман, изследователи в полето на мрежовите науки, използват изследванията за Закари за социални мрежи в своето изследване върху социални и биологични мрежи.²⁰ Налична е ясна връзка (социални мрежи/групи),

¹⁸ Пак там.

¹⁹ **Zachary**, Wayne. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, 1976, 33(4). <https://www.researchgate.net/publication/248519014> (04.09.2018).

²⁰ **Girvan**, Michelle et al. Community structure in social and biological networks. *PNAS*, 2002, 99(12), pp. 7821-7826. <http://www.pnas.org/content/99/12/7821> (04.09.2018).

но подходът на двете публикации е коренно различен. Изследването на Закари се базира изцяло върху методиката на социалните изследвания, докато Гирван и Нюман прилагат математичен модел, свързан по-тясно с информационните науки. Съответно може да направим заключението, че за пробуждането на „спяща красавица” не винаги е нужна директна връзка между нея и нейния принц (пробуждаща публикация). В случаи като гореописания пробуждането може да бъде основано на новооткритата приложимост на недооцененото изследване от изцяло различно научно поле. Днес в системата на Google Scholar изследването на Закари има акумулирани общо 3434 цитирания.²¹ Впечатляващо постижение за тясноспециализирано изследване, което се е намирало в състояние на неизвестност за почти десетилетия.

Интересно е също така да се отбележи, че процент (4%) от идентифицираните „спящи красавици” са публикувани в големи научни списания като Nature и Science. Това показва, че дори статии, публикувани в списания с много висок импакт фактор и съответно висока видимост, могат да останат незабелязани и недооценени.

Като цяло В-коефициентът е ефективна система за идентифициране и анализ на публикации от типа „спяща красавица”. Коефициентът е изграден по начин, който се съобразява с основните характеристики на тези публикации: брой първоначални цитирания, продължителност на съня и сила на пробуждането. Въпреки това тази система не е безпогрешна. Ке и неговите колеги отбелязват, че системата не взема предвид тенденциите за натрупване на цитирания, характерни за различните научни полета. Например една статия в полето на физиката има коренно различна тенденция за натрупване на цитирания от такова от полето на социалните науки. Въпреки това стойността на В-коефициента в отношение идентифицирането на недооценени научни публикации е безспорна.

Заключение

Днешната изследователска среда се характеризира с нейната конкурентна природа. Видимостта на изследванията на даден учен често се приравняват с това колко качествен е трудът му. Изследователски и финансиращи организации използват това като измерителна единица за това до каква степен дадено изследване е успешно, за да правят решения за това кои изследователи да биват повишавани и кои проекти да получават финансиране. В научноизследователската среда, пренаситена с огромно количество публикации, излизаци на чести интервали, желанието за унифицирана метрична единица, по която могат да се правят решения е разбираемо. В същото време, ако това продъл-

²¹ „An Information Flow Model for Conflict and Fission in Small Groups“ in Google Scholar: 1976–2018. <https://scholar.google.bg/scholar> (04.05.2018).

жи, рискуваме да създадем среда, където изследователските идеи и проекти се избират не заради тяхната научна стойност, а затова колко голям брой цитати те ще донесат на свързаните с изследването автори. Именно тук се намира стойността на „спящите красавици” и методи като В-коефициент. Трябва да може да покажем на изследователите, че дори ако техните публикации не намерят незабавно след своето публикуване отзвук в академичната среда, тяхната научна стойност не е неповратно изгубена завинаги.

ИЗПОЛЗВАНА ЛИТЕРАТУРА

- Van Noorden, Richard. Formula predicts research papers' future citations. *Nature*, 2013. <https://www.nature.com/news/formula-predicts-research-papers-future-citations-1.13881> (04.09.2018).
- Garfield, Eugene. Would Mendel's Work Have Been Ignored if Science Citation Index Was Available 100 Years Ago?. *Essays of an Information Scientist*, 1970, 1(2), pp. 69–70. <http://garfield.library.upenn.edu/essays/V1p069y1962-73.pdf> (04.09.2018).
- Garfield, Eugene. Premature discovery or delayed recognition: why?. *Essays of an Information Scientist*, 1980, 4(21), pp. 5-10. <http://garfield.library.upenn.edu/essays/v4p488y1979-80.pdf> (04.09.2018).
- van Raan, Anthony. Sleeping beauties in science. *Scientometrics*, 2004, 4, pp. 461–466. <https://pdfs.semanticscholar.org/9acc/dbe65099e8be88ef5553e8833e98ba28250e.pdf> (04.09.2018).
- Redner, Sidney. Citation Statistics from 110 years of *Physics Review*. *Physics Today*, 2005. <http://physics.bu.edu/~redner/pubs/pdf/PT.pdf> (04.09.2018).
- Einstein, Albert et al. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?. *Physics Review*, 1935, 44 (777). <https://journals.aps.org/> (04.09.2018).
- „Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935–1965. <https://scholar.google.bg/scholar> (04.09.2018).
- „Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935–1975. <https://scholar.google.bg/scholar> (04.09.2018).
- „Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” in Google Scholar: 1935–2018. <https://scholar.google.bg/scholar> (04.09.2018).
- Einstein-Podolsky-Rosen paradox observed in many-particle system for the first time. University of Basel, 2018. <https://www.unibas.ch/en/News-Events/News/Uni-Research/Einstein-Podolsky-Rosen-paradox.html> (04.09.2018).
- Garfield, Eugene. Citation indexes for science: a new dimension in documentation. *Science*, 1955, 122, pp. 108–111. <http://garfield.library.upenn.edu/papers/science1955.pdf> (04.09.2018).

- Ke, Qing et al. Defining and identifying Sleeping Beauties in science. PNAS, 2015, 112(24), pp. 7426–7431. <http://www.pnas.org/content/112/24/7426> (04.09.2018).
- Zachary, Wayne. An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research, 1976, 33(4). <https://www.researchgate.net/publication/248519014> (04.09.2018).
- Girvan, Michelle et al. Community structure in social and biological networks. PNAS, 2002, 99(12), pp. 7821–7826. <http://www.pnas.org/content/99/12/7821> (04.09.2018).
- „An Information Flow Model for Conflict and Fission in Small Groups“ in Google Scholar: 1976–2018. [https://scholar.google.bg/scholar\(04.05.2018\)](https://scholar.google.bg/scholar(04.05.2018)).