

EGC, une conférence qui supporte la diversité genrée?

Hélène de Ribaupierre *

* Cardiff University, Cardiff, CF24 3AA

deribaupierreH@cardiff.ac.uk,

<https://www.cardiff.ac.uk/people/view/543130-de-ribaupierre-helene>

Résumé. EGC supporte-t-elle la diversité genrée? L'objectif de cet article est de rapporter une démarche historique conduite sur l'histoire de la conférence EGC, dans une perspective de genre. Utilisant les données de publications de ces 20 dernières années, cet article analyse le rapport de genre entre les auteur·e·s des publications. L'article analyse le pourcentage de femmes publiant à EGC, et compare les résultats obtenus avec les conférences populaires auprès de ces auteur·e·s. L'analyse se fait à plusieurs niveaux (général, nombre d'auteur·e, premier·ère auteur·e, dernier·ère auteur·e). Ces analyses montrent que le pourcentage d'auteure d'EGC se situe dans la moyenne supérieure relativement aux autres conférences (27%); que les femmes publient moins en première auteure que les hommes, et qu'un certain niveau d'homophilie genrée existe dans les publications à auteur·e·s multiples. De plus les femmes semblent adopter une stratégie de publication différente de celle des hommes quant à la diversité des conférences dans lesquelles elles publient.

1 Introduction

La représentation des minorités dans les différents secteurs de la communauté scientifique est une question présente depuis maintenant quelques années. Malgré une prise de conscience des politiques et universitaires, l'augmentation du nombres de femmes scientifiques dans toutes les facettes de la vie académique est lente voire dans certains cas inexisteante (Danell et Hjerm, 2013). Par exemple, le pourcentage d'auteures dans les publications académiques est toujours plus bas que les hommes (Holman et al., 2018), même dans les domaines où les femmes sont mieux représentées qu'en informatique (de Ribaupierre et al., 2018); il en est de même pour les conférencier·ère·s invité·e·s (Débarre et al., 2018).

Il est reconnu que la dissémination du savoir académique se fait principalement par les publications scientifiques (West et al., 2013). Le domaine de l'informatique est un cas particulier car le poids des publications académiques comptant pour la reconnaissance des activités scientifiques sont publiées dans les journaux académiques, mais aussi dans les actes des conférences. Certaines conférences ont même atteint un niveau de réputation plus élevé que certains journaux. Par exemple, en Grande-Bretagne, REF¹ est un framework analysant la qualité des institutions universitaires et de recherche. Dans le cadre de ce framework, les publications sont

1. <https://www.ref.ac.uk/>

EGC, une conférence qui supporte la diversité genreé?

classifiées de une étoile à quatre étoiles. Une publication dans une conférence telle que IJCAI peut être classifiée comme une publications du plus haut niveau.

Les publications scientifiques sont non seulement la première source de dissémination des recherches scientifiques, mais elles sont aussi utilisées comme critères d'engagement et de promotion (Smith et al., 2013). Ainsi, les critères de qualité, le nombre de publications, le nombre de citations et, selon les domaines, la position de l'auteur·e dans les articles écrits à plusieurs sont pris en compte. En ce qui concerne la position de l'auteur·e, deux conventions ont été principalement adoptées et diffèrent entre domaines et sous-domaines ; elles ne sont pas toujours explicites. La première convention est de lister les noms d'auteur·e·s par ordre alphabétique, la deuxième est d'utiliser la première position dans l'ordre des auteur·e·s comme indication de la personne ayant écrit l'article, et la dernière position comme la plus senior. La recherche de (Le Berre, 2019) analyse la communauté francophone quand à ces pratiques, et observe que l'ordre semble moins important pour l'Intelligence Artificielle et les Systèmes d'information. Cependant, une analyse de l'ordre des auteur·e·s d'EGC montre que seulement 20% des publications utilisent un ordre alphabétique ; cette observation semble montrer que la position des auteur·e·s a une sémantique importante pour cette communauté.

La recherche menée par la présente auteure a porté sur les vingt dernières années de la conférence EGC, et a pris en compte la progression du nombre de femmes et le genre des auteur·e·s dans les publications multiples. Une comparaison avec les conférences les plus populaires auprès des auteur·e·s est proposée.

Un certain nombre de recherche concernant la représentation des femmes dans le domaine académique ont déjà été publiée, mais faire une revue de littérature complète dépasse le cadre de cet article.

2 Une analyse de la conférence EGC

Pour classifier les auteur·e·s en fonction de leur genre², des listes de prénoms incluant le sexe le plus probable du prénom ont été utilisées. Une approche automatique a été adoptée ce qui réduit la précision et le rappel relativement à une approche manuelle (Cabanac et al., 2016), cependant, cela permet d'étendre l'analyse à plus de conférences.

Étant donné que la communauté académique est internationale, une compilation de plusieurs listes est indispensable pour couvrir le plus de prénoms possible. Ces listes sont disponibles sur des sites Web tels que celui du gouvernement français³ ou (Raffo, 2016). Pour compléter ces listes, gender.api⁴ a été utilisé. Le taux de confiance retenu pour sélectionner le sexe de l'auteur·e est de 80%. Ce seuil réduit le nombre d'auteurs analysé et classifié : 88% pour ECG. Les prénoms écartés des listes peuvent l'être pour plusieurs raisons, dont la plus courante était un prénom mixte (Claude, Dominique, Camille,...).

Les metadatas des publications sont extraites du fichier XML de DBLP et sont les suivantes : titre, année de publication, nom des auteurs, nom de la conférence et la position de l'auteur·e. Suivant les domaines de recherche, cette dernière donnée est importante pour mieux comprendre la dynamique du rapport de genre dans les publications ; elle est souvent adoptée

2. dans cet article, une approche binaire du sexe a été adoptée

3. www.gard.gouv.fr

4. <https://gender-api.com>

comme critère de la qualité académique d'un·e chercheur·euse comme discuté dans l'introduction.

Pour extraire les conférences populaires auprès des auteur·e·s d'EGC, une liste de tous les auteur·e·s a été extraite. Cette liste est utilisée pour extraire toutes les conférences où un·e auteur·e d'EGC avait publié. Pour réduire cette liste, seules les conférences qui ont plus de 100 publications avec au moins un même nom qu'un·e auteur·e d'EGC ont été retenues. Les années analysées sont les mêmes pour toutes les conférences (tableau 1).

Nom des Conférence populaires	Nombre de publications	Langue
EDA	112	fr
RCIS	114	en
ICML	121	en
SEMWEB	123	en
CIKM	124	en
WWW	127	en
IJCAI	138	en
ICEIS	173	en
IC	192	fr
CORIA	201	fr
INFORSID	244	fr

TAB. 1 – *Conférences populaires auprès des auteur·e·s d'EGC, et nombre d'articles écrits par au moins un·e auteur·e d'EGC*

Distribution des sexes selon les auteur·e·s Depuis sa création en 2001, EGC présente un pourcentage de femmes auteures oscillant entre 19% et 33 %. La figure 1 montre la progression du nombres d'auteures au travers des 20 ans de cette conférence⁵. Le pourcentage de 1ère auteure est généralement plus faible que celui d'auteure, ce qui n'est pas le cas pour les hommes, qui publient légèrement plus en premier auteur. Pour les auteur·e·s en dernière position, ce pourcentage a augmenté durant ces vingt dernières années. Une régression linéaire a été calculée et montre que la projection sur les cinq prochaines années semble un peu meilleure pour les auteures que pour les premières auteures, et l'augmentation des dernières auteures semble être même un peu supérieure à celle des auteures.

La figure 2 illustre la moyenne des femmes auteures et premières auteures de ces 20 dernières années publiées dans d'autres conférences populaires auprès des auteur·e·s d'EGC. Les moyennes de CORIA, INFORSID, IC et RCIS sont un peu plus élevées que celle de EGC (27.4% des publications d'EGC comprennent des auteures femmes, et 26% des premières auteures). Ce pourcentage semble plus élevés pour les conférences francophones que pour les anglophones.

En moyenne, certaines conférences présentent plus de premières auteures que la moyenne d'auteures générale, ce qui peut s'expliquer par un nombre plus élevé de publications avec une auteure unique femme. Le pourcentage de publications avec auteur·e unique pour IC est de

5. publications analysées : 1602, nombre d'auteur·e·s 1203 (352 femmes, 851 hommes)

EGC, une conférence qui supporte la diversité genre?

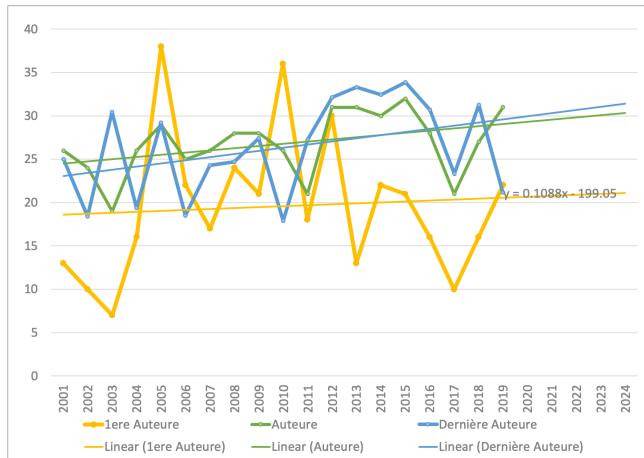


FIG. 1 – Pourcentage de femmes auteures et premières auteures, EGC

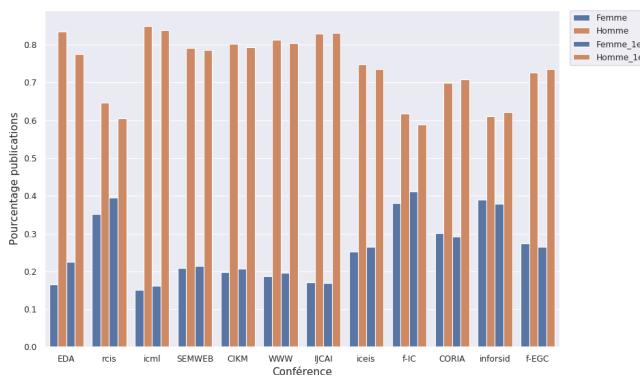


FIG. 2 – Pourcentage d'hommes et de femmes (en général) par conférence

9%, dont 31% écrites par des femmes; alors que pour EGC, 11% des publications sont écrites par un·e auteur·e unique, dont 21% écrites par des femmes.

Le pourcentage de femmes auteures dans les conférences a augmenté ces dernières années à un rythme différent selon la conférence (figure 3). EGC a augmenté de 26% à 27.4% en 20 ans. Une régression linéaire a été calculée, certaines conférences semblent avoir une augmentation plus rapide que d'autres (par exemple IC) et pourrait atteindre plus rapidement une parité dans le genre des auteur·e·s que d'autres (projection sur 5 ans). Une conférence présente un pourcentage de femmes en diminution au fil des années, il s'agit de RCIS.

Femmes auteures par nombre d'auteur·e·s Un grand nombre d'articles dans le domaine informatique sont écrits par plus d'un·e auteur·e. Pour EGC, le maximum d'auteur·e·s ana-

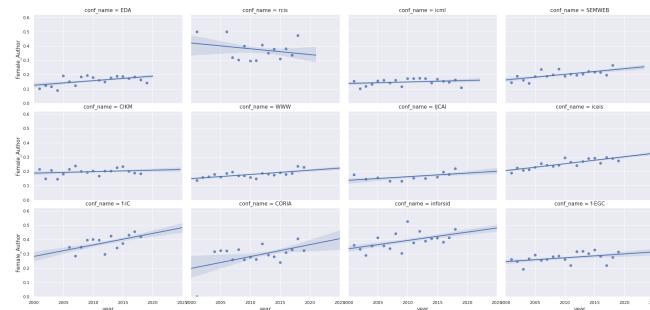


FIG. 3 – Pourcentage d'auteures par conférence par année

lysé est ici de 10 ; la médiane est de 3 auteur·e·s par article. Pour simplifier les tableaux et les analyses, les articles écrits par plus de 5 auteur·e·s ont été regroupés dans une même catégorie : 6 ou plus. Pour ECG (tableau 2), le pourcentage de publications écrites par des hommes seulement est plus élevé que celui des femmes ou celui d'un homme et une femme, ce qui semble indiquer une forte homophilie genrée. Plus le nombre d'auteur·e augmente, plus le niveau d'homophilie genré semble augmenter. On peut noter que le nombre d'articles écrits par quatre auteurs uniquement masculin est assez élevé, et que seulement 6% des articles à trois auteur·e·s sont écrits par des femmes uniquement contre 37% pour les hommes.

	Pourcentage de femmes dans la liste des auteurs				
Nombre d'auteur·e·s	100%	99-51%	50/50	0.1-49%	0%
2	0.08	-	0.39	-	0.53
3	0.06	0.17	-	0.39	0.37
4	0.0	0.07	0.23	0.30	0.41
5	0.0	0.21	-	0.51	0.29
6 ou plus	0.0	0.15	0.08	0.54	0.23

TAB. 2 – Diversité des auteur·e·s pour les publications à plus de deux auteur·e·s pour EGC

L'homophilie genrée semble présente dans les différentes conférences et quelques soit le nombre d'auteur·e·s (figure 4) ; IC, RCIS, CORIA et INFORSID sont les seules conférences où une parité de la co-création pour les articles écrit à deux auteur·e·s semble exister. Dans le cas des autres conférences, on observe une nette homophilie genrée en faveur des hommes. Par exemple, pour ICML, 75% des articles co-écrits par deux auteur·e·s sont des hommes. En comparaison avec les autres conférences, EGC se situe souvent dans la moyenne.

Nombre de publications par auteur·e·s par conférence, diversité des conférences par auteur·e·s Une part importante de la reconnaissance scientifique des scientifiques est de faire partie d'une ou plusieurs communautés ; pour ce faire, la/le scientifique peut adopter la straté-

EGC, une conférence qui supporte la diversité genre?

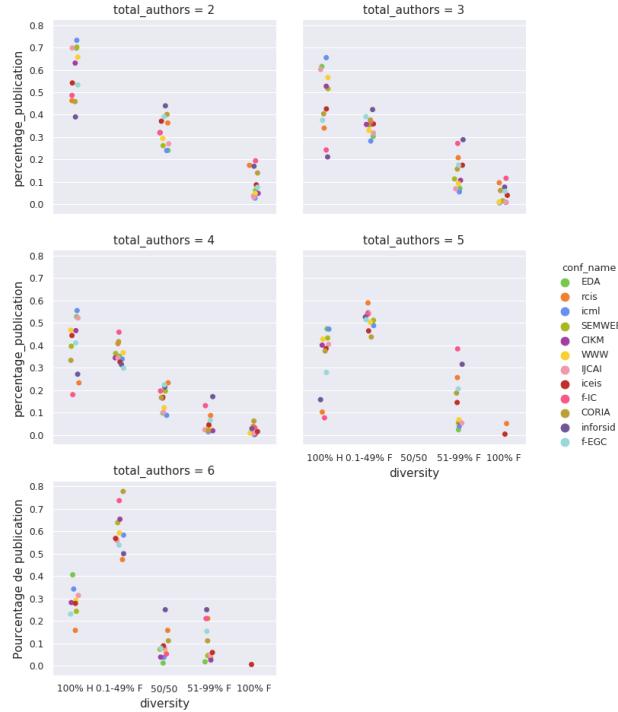


FIG. 4 – Diversité des auteurs : Publications à plus de deux auteur·e·s

gie de publier dans la même conférence sur plusieurs années. L'auteur le plus prolifique à EGC a publié 41 fois, contre 25 fois pour la première femme.

Il semblerait que le pourcentage de femmes publant plus d'une fois (re-publient) à EGC (5.4%) est plus faible que celui des hommes (13.5%). La conférence dans laquelle les femmes re-publient le plus est INFORSID (7.7% F, 9.9% H) suivie de près par IC (6.6% F, 10.9% H). SEMWEB est la conférence où les hommes re-publient le plus (4.4% F, 17.4%H). Les conférences présentant le pourcentage de femmes auteures généralement le plus faible, semblent aussi être celles où les femmes re-publient le moins.

Pour agrandir son réseau scientifique, et par la même occasion sa reconnaissance scientifique, la/le scientifique va essayer non seulement de publier dans la même conférence plusieurs fois, mais elle/il va aussi essayer de publier dans différentes conférences de manière à agrandir son réseau scientifique. Pour cette analyse la liste des conférences populaires a été étendue à 24 conférences (figure 5). Ces résultats montrent qu'il semble exister une différence dans les conférences choisies par les auteur·e·s d'EGC. Les auteures d'EGC semblent montrer une plus grande inclination pour les conférences CORIA, ECAI, EDA, EWS, ICDE, SEMWEB, SIGMOD, TITS, VLDB, EKAW, IC, ICEIS, INFORSID et RCIS ; alors que les auteurs semblent avoir une inclination plus grande pour AAAI, CIKM, COLT, IJCAI, NIPS, WWW, DAWAK, ICML et ICONIP. SIGMOD semble être la conférence qui attire autant les hommes que les femmes.

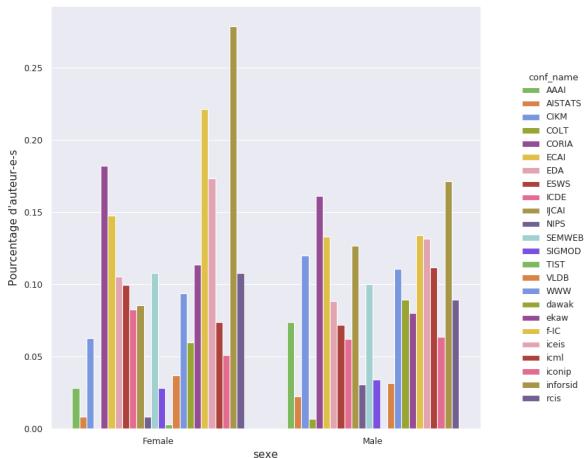


FIG. 5 – Pourcentage d'auteur-e-s d'EGC avec des publications dans les conférences similaires (pourcentage a été rapporté aux nombres d'auteur-e par genre)

3 Discussion et Conclusion

L'objectif de ce papier était de retracer les vingt années de la conférence EGC depuis sa création selon un point de vue genré. Cette recherche de nature historique a montré que les femmes partagent 27% de la co-création des publications d'EGC, ce qui la situe dans la moyenne supérieures relativement aux autres conférences. Ce taux de participation n'a pas beaucoup évolué durant cette période. Pour atteindre une égalité auteur-e pour EGC, il faudra attendre environ 50 ans, sur la base de la projection calculée par régression linéaire. Il est évident que cette égalité ne pourra se produire que si le nombre d'étudiantes en informatique augmente. Cette augmentation semble assez lente (Cabanac et al., 2016; de Ribaupierre et al., 2018).

La stabilisation des scientifiques dans certains postes académiques, pourrait expliquer que le nombre de femmes en dernière auteure augmente plus rapidement que le nombre de femmes en première auteure. Le pourcentage de co-auteures dans les articles semble plus faible pour EGC que pour d'autres conférences, et cette homophilie de genre semble augmenter en fonction du nombre d'auteur-e-s par article. On observe aussi dans EGC, comme dans d'autres conférences, que le nombre de publications par personne est plus élevé pour les hommes que pour femmes, ce n'est donc pas seulement une question de ratio femmes/hommes dans la communauté scientifique, mais aussi une question de publications par personne, les hommes publient plus que les femmes (West et al., 2013).

L'analyse des conférences populaires auprès des auteur-e-s d'EGC semble indiquer que les femmes et les hommes ne publient pas dans les mêmes conférences à l'exception de SEMWEB. Il serait intéressant d'analyser la thématique de ces conférences, et de voir s'il existe une corrélation entre ces thèmes, les champs de recherche dits "plus féminins", et les thématiques

EGC, une conférence qui supporte la diversité genreé?

des publications d'EGC.

Faute de pouvoir analyser exhaustivement le genre de tous les auteur.e.s d'EGC, certaines données peuvent manquer, mais le taux de précision de la classification était néanmoins de plus de 88%. Il serait intéressant dans un proche futur d'améliorer encore le rappel. Une prochaine recherche pourrait aussi analyser le taux de citation des auteur.e.s de EGC et de comparer le réseau de citations et le réseau de publications.

Références

- Cabanac, G., G. Hubert, H. D. Tran, C. Favre, et C. Labbé (2016). Un regard lexico-scientométrique sur le défi EGC 2016. In *16eme Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance (EGC 2016)*, Reims, France, pp. pp. 419–424.
- Danell, R. et M. Hjerm (2013). Career prospects for female university researchers have not improved. *Scientometrics* 94(3), 999–1006.
- de Ribaupierre, H., K. Jones, F. Loizides, et Y. Cherdantseva (2018). Towards gender equality in software engineering : the nsa approach. In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, pp. 10–13. ACM.
- Débarre, F., N. O. Rode, et L. V. Ugelvig (2018). Gender equity at scientific events. *Evolution Letters* 2(3), 148–158.
- Holman, L., D. Stuart-Fox, et C. E. Hauser (2018). The gender gap in science : How long until women are equally represented ? *PLOS Biology* 16(4), 1–20.
- Le Berre, D. (2019). societe-informatique-de-france/referentiel-pratiques-publication-2019.
- Raffo, J. (2016). Worldwide gender-name dictionary.
- Smith, K. M., E. Crookes, et P. A. Crookes (2013). Measuring research 'impact' for academic promotion : issues from the literature. *Journal of Higher Education Policy and Management* 35(4), 410–420.
- West, J. D., J. Jacquet, M. M. King, S. J. Correll, et C. T. Bergstrom (2013). The role of gender in scholarly authorship. *PLOS ONE* 8(7), 1–6.

Summary

Does EGC support gender diversity? The objective of this article is to report, from a gender perspective, an historical approach conducted on the EGC conference. Using data from the past 20 years, this article analyses the percentage of authorship in the EGC conference and other, related conferences. The analysis is conducted for the first author, the last author and in general. Results show that the percentage of women associated with publications in EGC (27%) is close to the high average of other conferences. Women seem to publish less as first author than men. Inbreeding Gender homophily seems to exist when the multi-authored publications are analysed, and grows with the number of authors. The publications strategy seems also to be different depending on the gender: men publish in different related conferences than women.