

1 We thank all reviewers for their insightful comments and suggestions, which will be incorporated into the revised
2 version. Our source code and the processed 950K USPTO-full dataset will be released to promote the related future
3 research if accepted. We first address a common concern about the concurrent work G2Gs[1].

4 The concurrent work G2Gs presents a similar two-step framework, while our method is more general and scalable.
5 G2Gs predicts at most one bond disconnection while our method can predict multiple bond disconnections, which is
6 more general. The scalability of our method is verified on USPTO-full while G2Gs does not scale to USPTO-full since
7 it can only cover 84.77% training reactions (Appendix Table 6). Besides, we formulate the reactant generation as a
8 seq2seq task by representing the molecule/synthon in SMILES, while G2Gs adopts graph representation and generates
9 reactants with the graph generation algorithm. Last but not least, our model can consider already generated reactants
10 when generating the next reactant (see L156 and L172 in our submission), while G2Gs independently generates multiple
11 reactants for each target, which goes against the nature of chemical reaction.

12 We keep muted about G2Gs before its conference version is available since we have some concerns about it. The major
13 concern is that the detailed composition of atom features is not given in their paper [1]. The minor concern is that they
14 failed to demonstrate the scalability on the USPTO-full. This discussion will be included into our revised version.

15 **Reviewer 1 Q1: How hyperparameters were selected, and how the validation dataset was used:** We find the
16 optimal hyperparameters with the best performance on the validation dataset. The model is robust to hyperparameters,
17 and we did not exhaustively search for hyperparameters. Both EGAT and RGN are trained for a fixed number of epochs
18 (L222-227), and results of the final EGAT model are reported and early stopping is unnecessary.

19 **Q2: EGAT network does not overcome the limitation of capturing only local information:** Capturing only local
20 information is a common limitation for MPNNs, and the EGAT belongs to MPNNs. We are not to design a new
21 architecture to overcome the limitation, but to mitigate it with a graph-level task predicting the number of disconnections.

22 **Q3: The synthon approach and semi-template rely on atom mapping:** Atom mapping is optional for our method.
23 The synthon approach can also work for reactions without provided atom mapping (L208-212). The EGAT performance
24 degrades only slightly when removing the semi-template indicator (L143-146).

25 **Q4: Could the authors clarify the clause “which seem to get stuck in an infinite loop” (L38):** Manual encoding
26 is infeasible for large reaction datasets. While template extraction algorithms rely on accurate atom mapping, which
27 requires expert rules. It comes back to the infeasibility issue. We will revise the description and language.

28 **Other questions:** The ground truth synthons are used for the evaluation in Table 2 (L264). L110: $K > 1$ does not
29 indicate multiple reactions with the same product are merged. K is the total number of reactions in training set, since
30 loss should be aggregated across the full training set. Semi-templates are extracted from training set as indicated at
31 L136. We will revise the discussion of template-based methods (L28, L34-35) as suggested to make it more clear.

32 **Reviewer 2 Q1: Compare performance on a different independently constructed dataset.** As far as we are con-
33 cerned, we are not aware of other large and public chemical reaction datasets except for USPTO-50K and USPTO-full.

34 **Q2: Whether the general strategy of splitting and conversion has been attempted:** This general strategy can be
35 found in the very early work [2]. We instantiate the strategy with deep learning models, which are novel and effective.

36 **Q3: Multi-step retrosynthesis and related prior work:** Monte Carlo tree search can be applied to the single-step
37 retrosynthesis recursively until reaching available molecules. We will include the suggested related work. Recent
38 template-free approaches formulate the retrosynthesis as a translation problem (L39-42), except for the G2Gs [1].

39 **Reviewer 3 Q1: How to obtain the label for the bond disconnection:** The label comes from the atom mapping
40 which it is provided by the USPTO datasets. Our method also works for datasets without atom mapping, since the
41 disconnection label can also be obtained using Maximum Common Substructure algorithm (L208-212).

42 **Q2: Semi-templates indicator:** Sorry for the ambiguity. Your understanding is right. Note that semi-template
43 subgraphs may contain strategic bonds as well as some neighboring bonds and atoms which provide important reaction
44 context. The detailed composition of semi-templates depends on the adopted reaction template extraction algorithm.

45 **Other questions:** We will make our statement more humble and practical as suggested. The suggested references
46 are closely related to our work and will be included in the revised version. It is easy and preferred to extract bond
47 disconnection labels from atom mapping. At the same time, our method also works for datasets without atom mapping.
48 Strictly speaking, it is not fully template-free because of the semi-template. However, the semi-template is not a must
49 option for EGAT, and its performance degrades only slightly when removing the semi-template indicator (L143-146).

50 [1] Shi, Chence, et al. "A Graph to Graphs Framework for Retrosynthesis Prediction." ICML, 2020.

51 [2] Pensak, David A., et al. "LHASA—Logic and Heuristics Applied to Synthetic Analysis." ACS, 1977. 1-32.