Kossinet's and Watts' paper seeks to answer the following research question: Does structural proximity or individual preference for similarity, or the dynamic interaction between both forces, account for homophily? As indicated in the introduction section, the homophily principle, is the observed tendency that we like to associate with people who are similar to each other, that is, spouses, colleagues, friends are more similar to each other than randomly chosen members of the population. As for structural proximity, or induced homophily, it is defined as homophily that arises as a consequence of the structural opportunities for interactions like in neighborhoods, while individual preference, or choice homophily, is defined as the active choices made by the individual to prefer engaging with people who are similar to that individual.

To test the hypothesis, they merged three sets of data: (1) the logs of e-mail interactions between 30,396 undergraduate and graduate students, faculty, and staff in a large U.S. university over one academic year, (2) a database of individual attributes, specifically, gender, age, status, field, year, and state (the last was converted to U.S./foreigner) (3) records of course registration. In particular, they used the first dataset to analyze the strength of interpersonal communication between two individuals, while they used the second dataset to analyze the aggregate similarity between each pair, from 0 and 6 across these dimensions. Their specific definitions of these dimensions like age, formal status, etc. are also indicated in appendix A. They also looked at only certain observations – they set the maximum time at which an interaction is assumed to contribute to the relationship or τ, as 60 and the sampling period δ, determining whether events separated in time will be treated as related with one another, as 1 day. From these parameters, Kossinet and Watts analysed specifically 210 sequential network snapshots within their data, which span the second half of the fall semester and the entire spring semester. Lastly, they used the third dataset as a measure of structural proximity. In particular, they used it to analyze explicit social foci, (and also used bulk email messages as an indicator of implicit social foci), where social foci is defined as the various groups, contexts, and activities which facilitate interpersonal interactions, like taking the same class. Their overall results not only show that homophily is displayed, but also show that the effects of homophily are strongly mitigated due to induced homophily, either because of shared mutual acquaintances or social foci, which suggest that structural proximity plays a more important role than individual choice in homophily.

Although they provide a very convincing argument, one potential critique is that they used only messages that were sent to a single recipient in the data cleaning process. Although this category accounts for a significant amount of all emails (82%), using this method disregards group interaction and may account only for a considerably small portion of social interactions; homophily may, in fact, exist and may even be more significant between a group of individuals, and not just between single pairs of individuals. As such, their conclusion that the origins of homophily is accounted by both individual choice and structural proximity (and more so by the latter), may not account for all cases of social interactions, and may only be specific to interactions between single pairs instead.

Additionally, in terms of internal validity, one can also argue that email logs between individuals in a university may not be an accurate indicator of the depth of social relationships. For example, emails between individuals in a university may be just be one-off and in a professional capacity, like between a pair of faculty staff in different departments, who are organizing driving arrangements to attend a one-off event. Kossinet and Watts do attempt to account for these short-term relationships, by limiting τ to 60 days so that their analysis will not include these one-off email bursts that are not accurate indicators of the formation of social relationships. That being said, they also acknowledge that they are also limited by their data; their dataset does not have texts of these email logs, for example, and running text analysis could provide a fuller picture on whether the communication within these emails are strictly

professional (and therefore are not relevant to their hypothesis) or are more personal and friendly, which indicate the formation of a social relationship.